Automated mining of the ALMA archive in the COSMOS field (A3COSMOS): I. Robust ALMA continuum photometry catalogs and stellar mass and star formation properties for ~700 galaxies at z = 0.5–6

Article (Accepted Version)

Draft version September 13, 2019
Typeset using LaTeX **twocolumn** style in AASTeX62

Automated Mining of the ALMA Archive in the COSMOS Field (A^{3}COSMOS): I. Robust ALMA Continuum Photometry Catalogs and Stellar Mass and Star Formation Properties for ∼700 Galaxies at $z = 0.5$–6

Daizhong Liu,[1] P. Lang,[1] B. Magnelli,[2] E. Schinnerer,[1] S. Leslie,[1] Y. Fudamoto,[3] M. Bondi,[4] B. Groves,[5] E. Jiménez-Andrade,[2] K. Harrington,[2] A. Karim,[2] P. A. Oesch,[3] M. Sargent,[6] E. Vardoulaki,[2] T. Bădescu,[2] L. Moser,[2] F. Bertoldi,[2] A. Battisti,[5] E. da Cunha,[5] J. Zavala,[7] M. Vaccari,[8,9] I. Davidzon,[10] D. Riechers,[11] and M. Aravena[12]

[1] *Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany*
[2] *Argelander-Institut für Astronomie, Universität Bonn, Auf dem Hügel 71, D-53121 Bonn, Germany*
[3] *Department of Astronomy, Université de Genève, Chemin des Maillettes 51, 1290 Versoix, Switzerland*
[4] *INAF - Istituto di Radioastronomia, Via Gobetti 101, 40129, Bologna, Italy*
[5] *Research School of Astronomy and Astrophysics, Australian National University, Canberra ACT, 2611, Australia*
[6] *Astronomy Centre, Department of Physics and Astronomy, University of Sussex, Brighton BN1 9QH, UK*
[7] *The University of Texas at Austin, 2515 Speedway Boulevard Stop C1400, Austin, TX 78712, USA*
[8] *Department of Physics and Astronomy, University of the Western Cape, Robert Sobukwe Road, 7535 Bellville, Cape Town, South Africa*
[9] *INAF - Istituto di Radioastronomia, via Gobetti 101, 40129 Bologna, Italy*
[10] *IPAC, Mail Code 314-6, California Institute of Technology, 1200 East California Boulevard, Pasadena, CA 91125, USA*
[11] *Department of Astronomy, Cornell University, Space Sciences Building, Ithaca, NY 14853, USA*
[12] *Núcleo de Astronomía, Facultad de Ingeniería, Universidad Diego Portales, Av. Ejército 441, Santiago, Chile*

(Revised September 13, 2019)

ABSTRACT

The rich information on (sub-)millimeter dust continuum emission from distant galaxies in the public Atacama Large Millimeter/submillimeter Array (ALMA) archive is contained in thousands of inhomogeneous observations from individual PI-led programs. To increase the usability of these data for studies deepening our understanding of galaxy evolution, we have developed automated mining pipelines for the ALMA archive in the COSMOS field (A3COSMOS) which efficiently exploit the available information for large numbers of galaxies across cosmic time, and keep the data products in sync with the increasing public ALMA archive: (a) a dedicated ALMA continuum imaging pipeline; (b) two complementary photometry pipelines for both blind source extraction and prior source fitting; (c) a counterpart association pipeline utilizing the multi-wavelength data available (including quality assessment based on machine-learning techniques); (d) an assessment of potential (sub-)mm line contribution to the measured ALMA continuum; and (e) extensive simulations to provide statistical corrections to biases and uncertainties in the ALMA continuum measurements. Application of these tools yields photometry catalogs with ∼ 1000 (sub-)mm detections (spurious fraction ∼ $8-12\%$) from over 1500 individual ALMA continuum images. Combined with ancillary photometric and redshift catalogs and the above quality assessments, we provide robust information on redshift, stellar mass and star formation rate for ∼700 galaxies at redshifts 0.5-6 in the COSMOS field (with undetermined selection function). The ALMA photometric measurements and galaxy properties are released publicly within our blind-extraction, prior-fitting and galaxy property catalogs, plus the images. These products will be updated on a regular basis in the future.

Keywords: galaxies: photometry — galaxies: star formation — galaxies: evolution — galaxies: ISM — submillimeter: galaxies — techniques: photometric

1. INTRODUCTION

The interstellar medium (ISM) is the raw material in galaxies out of which stars form. It plays a fundamental role when reconstructing the Universe through cosmological simulations. In galaxies harboring intensive star

dzliu@mpia.de

formation, cold neutral gas is the main component of the ISM dominating its mass; and a significant fraction of this cold gas is in the molecular phase (e.g., Walter et al. 2008; Leroy et al. 2008; Bigiel et al. 2008). Over the past four decades, molecular gas has been observedmainly via the Carbon Monoxide (CO) rotational transition lines in the rest-frame millimeter (mm) wavelengths (which are the most feasible observable tracers of molecular gas; e.g., see review by Solomon & Vanden Bout 2005; Carilli & Walter 2013). However, CO observations at high redshift ($z > 1$) mostly target the brightest sub-millimeter galaxies (SMGs; e.g., review by Blain et al. 2002) and quasi-stellar objects (QSOs). These objects are the most extreme cases and not representative of the more numerous, less starbursty galaxies, i.e., the star-forming galaxies that follow a tight main sequence (MS) in the stellar mass–SFR plane (e.g. Brinchmann et al. 2004; Noeske et al. 2007; Elbaz et al. 2007; Daddi et al. 2007). Observing CO in a large number (e.g., a few hundred) of main-sequence galaxies at $z > 1$ (hence probing the ISM evolution) is in practise very time-consuming even with the most advanced facility, the Atacama Large Millimeter/submillimeter Array (ALMA), as firstly all galaxies are required to have a spectroscopic redshift in advance, secondly sufficient sensitivity is needed to detect the line within a small spectral bandwidth (typically $\sim 300 - 500$ km/s), and thirdly the galaxy sample should cover enough parameter space in the main sequence plane.

In recent years, a much more efficient approach — carrying out broad-band dust continuum observations to infer ISM — has been established. With the use of all the bandwidth of the receiver, usually much short integration time are needed for high-redshift galaxies than observing (sub)mm emission lines (see also Carpenter et al. 2019). Then the cold gas mass can be inferred either using the gas-to-dust mass ratio, $\delta_{\rm GDR}$, which has been reasonably characterized as a function of gas phase metallicity (e.g., Santini et al. 2010; Leroy et al. 2011; Magdis et al. 2011, 2012; Magnelli et al. 2012; Bolatto et al. 2013; Rémy-Ruyer et al. 2014; Tan et al. 2014; Coogan et al. 2019), or with the ratio between gas mass and dust continuum luminosity at Rayleigh-Jeans tail wavelengths (e.g., rest-frame 250–850 μm) which has been calibrated with rich observations (e.g., Scoville et al. 2014; Groves et al. 2015; Hughes et al. 2017; Bertemes et al. 2018; Saintonge et al. 2018). The use of dust continuum observations to systematically survey the ISM content in hundreds of high redshift galaxies is already proved to be fruitful (e.g., Schinnerer et al. 2016; Scoville et al. 2016, 2017).

Meanwhile, the continuously growing ALMA public archive offers a great opportunity of studying very large samples of high-redshift galaxies. The ALMA public archive consists of thousands of observations of high-redshift galaxies within deep fields led by individual Principle Investigator (PI) programs. Although ALMA has a small field of view, e.g., $\sim$0.5′ in diameter (primary beam FWHM) in Band 6, accumulating archival data compensates for this shortcoming and leads to several hundred arcmin^2 area. Comparing to contiguous deep field surveys with ALMA (e.g., Hatsukade et al. 2011, 2016, 2018; Carniani et al. 2015; Walter et al. 2016; Aravena et al. 2016; Dunlop et al. 2017; Umehata et al. 2017, 2018; Franco et al. 2018), the discreteness of field of views makes the sample selection bias and cosmic comoving volume very unpredictable, but it also leads to a sample with large varieties in galaxy properties, which thus provides crucial constraints on galaxy ISM and star formation scaling relations and analytic evolution prescriptions (e.g., Scoville et al. 2017; Tacconi et al. 2018). Moreover, the ALMA archive also serves as a powerful test bed for automated pipelines as in this work and for future large facilities.

Several recent studies have already been exploring the full ALMA archive (e.g. Scoville et al. 2017; Fujimoto et al. 2017; Zavala et al. 2018), however, the photometric methods (including aperture photometry, peak pixel analysis, *uv*-plane fitting, etc.) and sample selection can significantly differ between different authors. Furthermore, none of these studies have statistically evaluated the effects of applying different photometric methods to ALMA images, especially for large numbers of ALMA images with widely varying sensitivity and synthesized beam properties. Consequently, noticeable discrepancies on the cosmic evolution of the ISM are present among the aforementioned studies. In order to understand how potential biases of the photometric methods and gas mass calibration affect the outcome of ISM evolution studies more dedicated efforts are required to exploit the public ALMA archive.

In this work, we present automated pipelines for "mining" the public ALMA archive in the COSMOS field (Scoville et al. 2007) (hereafter referred to as the "A^{3}COSMOS" (Automated mining of the ALMA Archive in COSMOS) project [1]. This work provides the foundation for a systematic exploitation of the (sub-)mm continuum as a proxy for cold dust and gas for a diverse and large sample of high-redshift galaxies. The resulting catalog of galaxies with (sub-)mm continuum detections

[1] https://sites.google.com/view/a3cosmos

can be used to, e.g., study the cosmic evolution of the gas fraction and gas depletion time (paper II; D. Liu et al. in prep.).

We present our workflow from the raw public ALMA data to the two robust photometric catalogs in Fig. 1, which corresponds to Sects. 2 to 3 of the paper. We first describe our ALMA data reduction and continuum imaging procedures in Sect. 2.1, then we present the blind source extraction in Sect. 2.2, prior catalog compilation in Sect. 2.3, and prior source fitting in Sect. 2.4. Sect. 3 includes our extensive Monte Carlo simulations and analyses to verify our photometry. Sect. 4 describes how we combine the two photometry catalogs, remove spurious sources and build a final well-characterized galaxy catalog (a workflow for these sub-steps is presented at the beginning of Sect. 4) Finally, the resulting catalogs are described in Sect. 5, while we summarize the paper in Sect. 6.

Throughout the paper, we adopt a flat ΛCDM cosmology with $H_0 = 70\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$, $\Omega_\mathrm{M} = 0.3$, $\Omega_0 = 0.7$, and a Chabrier (2003) initial mass function (IMF).

2. DATA AND PHOTOMETRY

The public ALMA archive is growing rapidly through PI-led observations. These observations mainly focus on targeted scientific objectives (sources) which are usually at the phase center of each ALMA pointing. However, with ALMA's unprecedented sensitivity, and benefiting from the negative-K correction at mm wavelengths (e.g., review by Blain et al. 2002; Casey et al. 2014), further (sub-)mm galaxies can serendipitously appear in any ALMA pointing. Such sources have a sizable chance for detection when their position falls within about twice the primary beam area [2] of the corresponding ALMA pointing (i.e., with a primary beam attenuation ≥ 0.2).

Here, we conduct a systematic effort to exploit these observational data. We limit our selection to within the COSMOS field (R.A. = 10:00:28.6, Dec. = +02:12:21.0, J2000; Scoville et al. 2007), because it is one of the deep fields with the richest, deepest multi-wavelength datasets, and there are numerous PI-led ALMA observations within its large area of 2 deg^2 (compared to the Great Observatories Origins Deep Survey (GOODS) North and South fields with only 160 arcmin^2 (0.044 deg^2) each). We include all the available ALMA data in COSMOS regardless of the ALMA Bands used (but excluded very long baseline data with a synthesized beam $< 0.1''$; see Sect. 2.1; and the only one mosaic project on AzTEC-3 protocluster).

COSMOS has extensive imaging datasets covering all accessible wavelength ranges: X-ray (Elvis et al. 2009; Civano et al. 2012, 2016; Marchesi et al. 2016), UV (Zamojski et al. 2007), optical (Leauthaud et al. 2007; Capak et al. 2007; Taniguchi et al. 2007, 2015), near-IR (McCracken et al. 2010, 2012), mid-IR (Sanders et al. 2007, Le Floc'h et al. 2009), far-IR (Lutz et al. 2011; Oliver et al. 2012), sub-mm (Geach et al. 2017), mm (Bertoldi et al. 2007; Aretxaga et al. 2011) and radio (Schinnerer et al. 2010; Smolčić et al. 2017). The depths of the X-ray, UV, optical and near-IR datasets are listed in Laigle et al. (2016), and the depths of mid-to-far-IR, (sub-)mm and radio datasets are summarized in Jin et al. (2018).

Photometric redshifts have been obtained for $\sim 1.1 \times 10^6$ galaxies through optical to near-IR spectral energy distribution (SED) fitting by Muzzin et al. (2013), Ilbert et al. (2013), Laigle et al. (2016), Davidzon et al. (2017) and Delvecchio et al. (2017), and through optical to mm/radio SED fitting by Jin et al. (2018).

Spectroscopic redshifts also exist for $\sim 7.1 \times 10^4$ galaxies, from the latest compilation by M. Salvato et al. (version Sept. 1st, 2017; available internally in the COSMOS collaboration), which includes almost all spectroscopic observations in the COSMOS field: Lilly et al. (2007, 2009; zCOSMOS Survey; with VLT/VIMOS); Fu et al. (2010; with Spitzer/IRS); Casey et al. (2012, 2017; with Keck II/DEIMOS); Comparat et al. (2015; with VLT/FORS2); Le Fèvre et al. (2015) and Tasca et al. (2017) (VUDS Survey; with VLT/VMOS); Hasinger et al. (2018; with Keck II/DEIMOS); Kriek et al. (2015; MOSDEF Surey; with Keck I/MOSFIRE); Marsan et al. (2017; with Keck II/NIRSPEC); Masters et al. (2017; with Keck II/DEIMOS); Nanayakkara et al. (2016; with Keck I/MOSFIRE); Silverman et al. (2015a; FMOS-COSMOS Survey; with Subaru/FMOS); van der Wel et al. (2016; LEGA-C Surey; with VLT/VMOS); Yun et al. (2015; with LMT/RSR) (listed only references whose spectroscopic redshifts are used in this work).

We show the pointings of all public ALMA data for the COSMOS field as of Jan. 2nd, 2018 in Fig. 2, overlaid on the *Herschel Space Observatory* (Pilbratt et al. 2010) Photodetector Array Camera and Spectrometer (PACS; Poglitsch et al. 2010) 100μm image. All the pointings processed for catalogs presented here are shown in green, and data which will be processed in our next release are shown in magenta, which includes ALMA data becoming public before August 1$^{\mathrm{st}}$, 2018. Circle size

[2] Here the primary beam area means the area enclosed in a circle with a radius equaling the primary beam's full width at half maximum (FWHM).

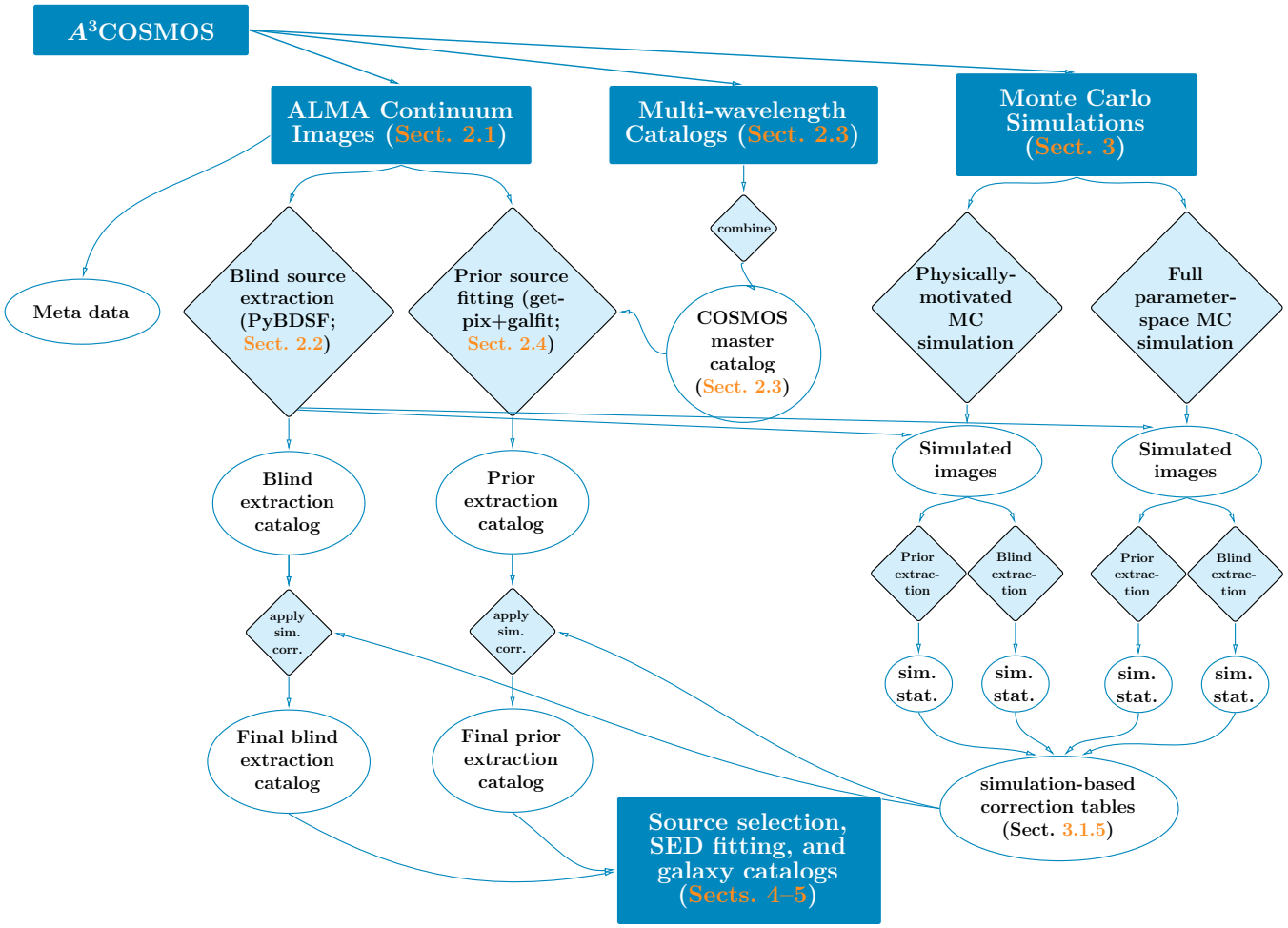

Fig. 1. Workflow of our automated mining of the ALMA archive in the COSMOS field (A^{3}COSMOS), which corresponds to Sects. 2 to 3 of this paper. **(Left branch:)** Starting from Sect. 2.1, we create ALMA continuum images from raw ALMA data which are obtained from the full public ALMA archive data for COSMOS. Next two photometric pipelines are used: (a) blind source extraction (see Sect. 2.2) and (b) prior source fitting (see Sect. 2.4) utilizing the COSMOS master catalog compiled beforehand as described in Sect. 2.3 **(middle branch)**. **(Right branch:)** Two Monte Carlo simulation pipelines with largely different prior assumptions are employed to verify the two photometric methods, and to provide flux bias correction and flux error estimation for the two photometric catalogs (see Sect. 3, and Appx. C for details). After verification, the two photometric catalogs are combined to build a galaxy catalog with physical properties as described in Sect. 4.

represents the primary beam [3]. The sum of primary beam area of these observations reaches 164 arcmin^2 as of Jan. 2nd, 2018, and will reach 280 arcmin^2 in our next release. Some pointings overlap because they are observed at different frequencies or with different spatial resolution. The overlapped area of all pointings is about 12%. Thus, even considering the non-overlapped primary beam area, the current data already reaches a spatial coverage similar to the area of the GOODS fields, and is much larger than any existing contiguous ALMA deep field survey (e.g., Dunlop et al. 2017, 4.5 arcmin^2 with 1σ $\sim$35 μJy/beam; Franco et al. 2018, 69 arcmin^2 with 1σ $\sim$0.18 mJy/beam).

In Fig. 3, we compare the depth and areal coverage of the ALMA archival data in COSMOS at ALMA Band 6 and 7 to the selected existing contiguous ALMA continuum deep fields: Aravena et al. (2016; see also Walter et al. 2016); Dunlop et al. (2017); and Franco et al. (2018; PI: D. Elbaz). Other ALMA deep fields (e.g., Hatsukade et al. 2016; Umehata et al. 2017) have similar properties and are therefore not shown. We compute the depth of each ALMA image by converting its rms

[3] Primary beam full widths at half maximum (FWHM) are computed according to https://www.iram.fr/IRAMFR/ARC/documents/cycle3/alma-technical-handbook.pdf, Eq. (3.4).

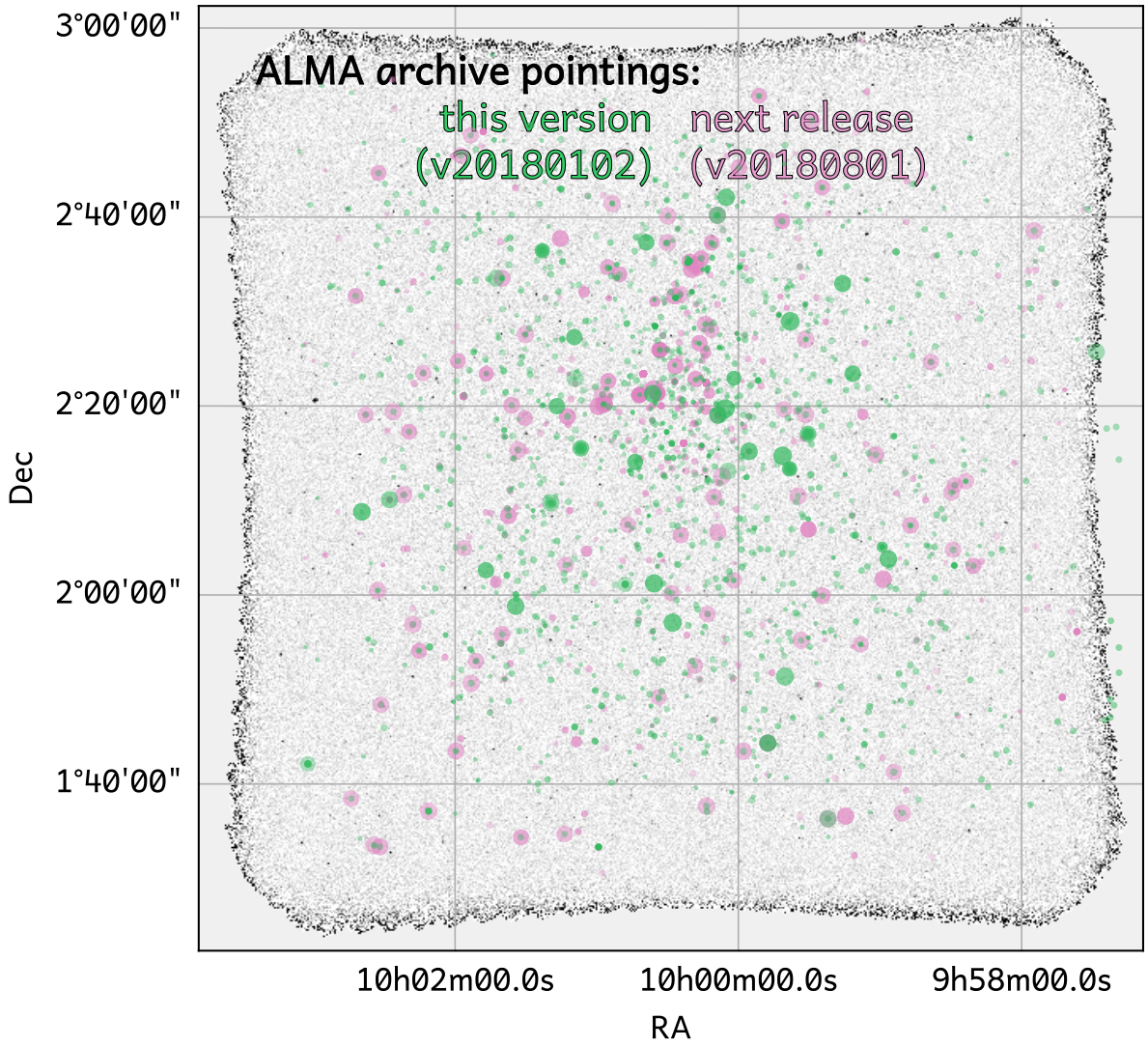

Fig. 2. ALMA pointings in the COSMOS field that are publicly accessible. Green and magenta circles represent ALMA pointings which became public before Jan. 2nd, 2018 and Aug. 1st, 2018 respectively. Circle sizes represent the FWHM of the ALMA 12m antennas' primary beam, and the shading reflects the on-source integration time (dark referring to longer integration times). The background image is the *Herschel* PACS 100 μm data from the PACS Evolutionary Probe survey (PEP; Lutz et al. 2011).

noise to an equivalent flux at observed-frame 1.1 mm assuming a modified blackbody with $\beta = 1.8$. The green (orange) curve represents the cumulative area of version 20180102 (version 20180801) ALMA images reaching a given sensitivity. Given that the ALMA archival data in the COSMOS field alone covers a larger cumulative area at all sensitivities, a systematic mining of the ALMA archive is strongly motivated. Given the inhomogeneous science goals of the individual PI-led projects, the resulting catalogs will not have a well characterized selection function and are not complete per se (see Sect. 4.7 for a discussion of the properties and completeness of the final galaxy catalog.)

In the following sections, we describe the reduction and processing of ALMA raw data into image products (in Sect. 2.1), and the photometric methods used (in Sects. 2.2 to 2.4). We employ two complementary photometric methods: blind source extraction and prior source fitting to obtain source flux densities and sizes from the ALMA continuum images. A comparison of the two methods and further technical assessments are presented in Sects. 2.5 to 2.7.

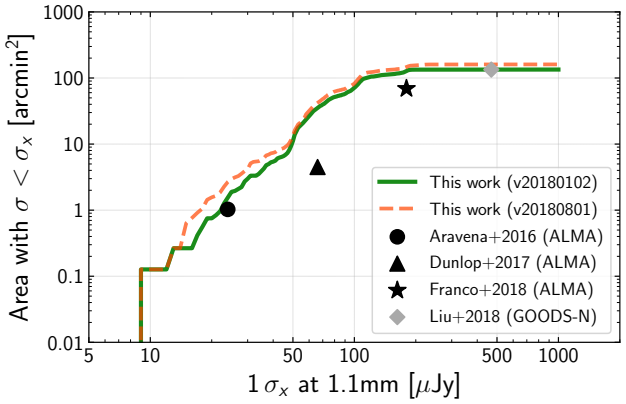

Fig. 3. (This figure has been revised.) Accumulated areal coverage of the public ALMA archival pointings at ALMA Band 6 and 7 used in this work as a function of the effective 1.1 mm 1-σ sensitivity (i.e., rms of pixel noise converted to the 1-σ sensitivity at 1.1 mm assuming a modified blackbody with $\beta = 1.8$; pixel size varies and is about 0.2× the beam size of each ALMA cleaned image). The orange line represents our next data release (corresponding to the magenta circles in Fig. 2). The area and sensitivity of three representative contiguous ALMA deep field surveys from (Aravena et al. 2016; see also Walter et al. 2016), Dunlop et al. (2017) and Franco et al. (2018) are shown as black symbols for comparison, as well as the 1-σ rms of super-deblended SCUBA2 photometry in the GOODS-North deep field from Liu et al. (2018) (gray diamond).

2.1. *ALMA Continuum Images*

We start by querying the ALMA archive with the `Python` package `astroquery` Ginsburg et al. (2019), retrieving all projects publicly available within a search radius of 2 degrees centered on the COSMOS field. These datasets are calibrated with the Common Astronomy Software Applications (CASA; McMullin et al. 2007) using the `scriptForPI.py` scripts provided by the Joint ALMA Observatory together with the archived raw data.

Calibrated visibilities are imaged and "cleaned" — i.e., deconvolved with the "dirty" beam — with the CASA imaging pipeline version `4.7.2`. With this systematic approach, we aim at obtaining data products as homogeneous as possible and also with maximized sensitivities. The pipeline is operated in "continuum" + "automatic" mode, leaving all but the weight parameters (set to "Briggs" with `robust=2`) to their default values. In this mode, the spectral windows (SpWs) of each target are aggregated into a single continuum image calculated at the central frequencies of these SpWs using the multi-frequency synthesis (`MFS`) algorithm with `nterms=2`. The parameters controlling the deconvolution process (i.e., masked pixels, maximum number of iterations and stopping threshold) are automatically and homogeneously set by the pipeline based on the noise properties and dynamic range of the "dirty" images (i.e., before deconvolution). The output images sample the synthesized beam with 5 pixels and are masked where the primary beam attenuation < 0.2. In case of obvious image artifacts in the cleaned images (as found by visual inspection, $< 10\%$), we re-run the CASA imaging pipeline flagging corrupted baselines and/or adopting `robust=0.5`.

The imaging pipeline uses masks to identify regions of bright emission prior to cleaning and the stopping criterion is set to a S/N=4. Given this approach combined with the sparseness of high S/N>15 sub-mm sources in our catalog, we do not expect any overcleaning resulting in artificially low rms noise. Given the large number of antennas in the 12m array, the instantaneous dirty beam for a short integration (∼30 seconds) as used for many programs has very low sidelobes, the presence of imaging artifacts is also minimized. The robustness of our cleaning process is also supported by our comparison of image-plane to uv-plane photometry (Sect. 2.6).

The CASA imaging pipeline could unfortunately not be run for a few of our projects (mostly from Cycle 0) due to backwards compatibility issues. These projects are thus imaged with the CASA task `clean` with input parameters manually set using a similar imaging and "cleaning" strategy as the CASA imaging pipeline, e.g., "Briggs" weighting with `robust=2`, sampling of the synthesized beam by ∼5 pixels and masking based on the noise of the dirty image.

As a test, we measure the $1\,\sigma$ sensitivities (rms noise) of our images ($\sigma_{\rm A^3}$) and compare with those measured in the continuum images available in the ALMA archive which are produced during the phase 2 of the ALMA Quality Assessment (QA2; $\sigma_{\rm QA2}$). $\sim 60\%$ of our images have QA2-based continuum images, while the remaining $\sim$ 40% were not imaged during the QA2 mostly because they are part of the scheduling blocks (SBs) with multiple targets, and the quality assessment was performed by imaging only a few of them. Consequently, for most projects, at least one QA2-based continuum image is available to perform our test. The $(\sigma_{\rm A^3} - \sigma_{\rm QA2})/\sigma_{\rm A^3}$ follows a Gaussian distribution centered at ~ -0.1 with a dispersion of ~ 0.17 (Fig. 4). Our images have ∼10% better sensitivities than those from the ALMA archive because they are produced with `robust=2` – i.e., favoring sensitivity over spatial resolution – while most QA2 analyses are performed with `robust=0.5`. We find no outliers with large positive values (e.g., $(\sigma_{\rm A^3} - \sigma_{\rm QA2})/\sigma_{\rm A^3} > 0.6$), as images with obvious artifacts were spotted by our visual inspection and already re-imaged. Finally, we find few outliers with

$(\sigma_{\rm A^3COSMOS} - \sigma_{\rm QA2})/\sigma_{\rm QA2} < -0.6$. We systematically checked these images and found that all of them correspond to projects in which the QA2-based analysis was performed with low `robust` values (i.e., < 0) and/or using only a fraction of the SpWs available. All these comparisons demonstrate the reliability of the ALMA imaging pipeline and thus of our image products.

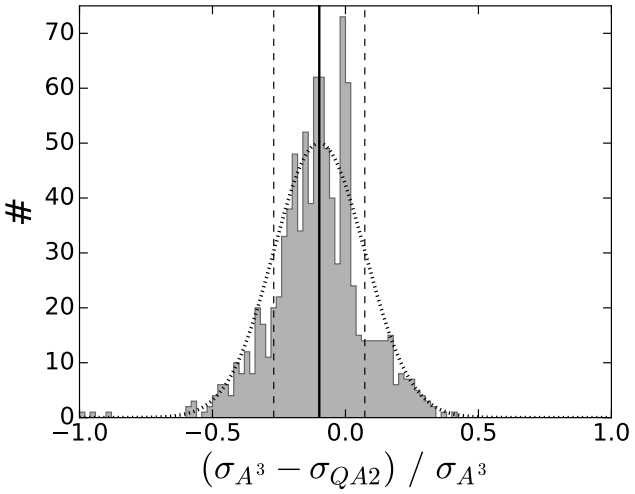

Fig. 4. Comparison of the $1-\sigma$ sensitivities of our images (i.e., $\sigma_{\rm A^3}$) to those measured in continuum maps from the ALMA archive (i.e., $\sigma_{\rm QA2}$). The dark histogram shows the $(\sigma_{\rm A^3} - \sigma_{\rm QA2})/\sigma_{\rm A^3}$ distribution, while the vertical continuous and dashed lines represent its mean (~ -0.1) and dispersion (~ 0.17). A Gaussian distribution with similar characteristics is shown as a dotted line.

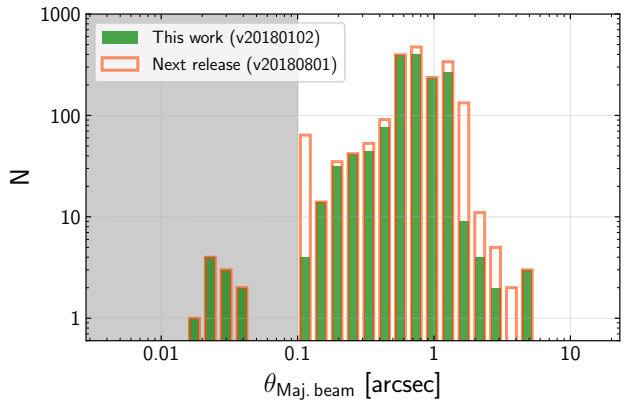

Fig. 5. Beam size distributions of two versions of A^{3}COSMOS data sets. The beam size is defined as the FWHM along the major axis of ALMA data's synthesized beam ($\theta_{\rm Maj.,\,beam}$) after interferometric "cleaning" process. The gray-shaded area indicates $\theta_{\rm Maj.,\,beam} < 0.1''$, for which data was discarded due to too high spatial resolution as discussed at the end of Sect. 2.1.

The data products released here include all "clean" continuum images corrected and uncorrected for primary beam attenuation (PBA)[4]. Note that although the aggregation of all SpWs available for a given target optimizes the sensitivities of our continuum images, it does not consider any possible line contamination. In Sect. 4.5 we will further describe an effective approach to addressing potential line contamination.

Furthermore, in Fig. 5, we show the distribution of the angular resolution of the ALMA data, as represented by the major axis FWHM of ALMA data's synthesized beam $\theta_{\rm Maj,\,beam}$. Of current images, $\lesssim 1\%$ have a very high angular resolution, i.e., $\theta_{\rm Maj,\,beam} < 0.1''$. These images represent a more challenging case for our source extraction because our blind and prior source extraction methods are all optimized for only marginally resolved sources, while sources in the very high resolution images usually are significantly resolved (e.g., the ratio of source to beam area $\gtrsim 10$). Also note that the large number of independent beams within these images ($\propto$ FoV / $\theta^2_{\rm Maj,\,beam}$) statistically translate into a significant contamination of "spurious" sources to our photometry catalog (even using a conservative S/N ~ 5 cut; see Sect. 2.8). Therefore these $\lesssim 1\%$ very high resolution ($\theta_{\rm Maj,\,beam} < 0.1''$) images are excluded from our analysis.

Currently, data for the same source taken at the same frequency arising from different projects is not combined.

The breakdown of the number of objects detected, the expected number of false objects, the area, median depth and resolution as a function of observing band are provided in Table 1.

2.2. *Blind Source Extraction*

We perform the blind source extraction on our "cleaned" ALMA continuum images. We use the primary-beam-attenuation-uncorrected images because they have the advantage of a constant noise across the field of view, and thus source extraction can be run with uniform parameters across them. The primary beam attenuation corrections are applied after the photometry steps.

We use the PYTHON BLOB DETECTOR AND SOURCE FINDER (a.k.a. PYBDSM or PYBDSF; hereafter PYBDSF; Mohan & Rafferty 2015)[5] to find sources

[4] The PBA corrections are due to the non-uniform sensitivity within the Gaussian-approximated FWHM of the primary beam for each antenna.

[5] PYBDSF documentation: http://www.astron.nl/citt/pybdsf/index.html; and its source code: https://github.com/lofar-astron/PyBDSF.

Table 1. Information per ALMA Band

Info Type	Band 3	Band 4	Band 5	Band 6	Band 7	Band 8	Band 9
Number of Images	34	6	2	633	857	1	1
Sum Beam Area [arcmin^2]	26.639	2.294	0.329	79.511	54.729	0.044	0.016
Mean Beam Size [arcsec]	2.164	1.098	1.548	1.202	0.772	0.526	0.305
Mean RMS Noise [mJy/beam]	0.039	0.025	0.090	0.077	0.160	0.034	1.757
PYBDSF S/N$_{\rm peak} > 5.40$	24	5	3	371	524	1	2
GALFIT S/N$_{\rm peak} > 4.35$ [a]	20 (7)	10 (7)	2 (2)	452 (342)	553 (461)	1 (1)	1 (1)

[a] The number in the brackets corresponds to the sources that passed our quality assessments from Sect. 4.1 to Sect. 4.4. Based on the spurious fraction analysis in Sect. 2.8, we expect about 8% spurious sources in total for the PYBDSF selection and ∼12% for the GALFIT selection.
[b] The areas are the sum of primary beam circular area only.

blindly and extract their flux and size information. First, the code identifies "islands" of emission, i.e., with the peak pixel emission above 4 times the rms noise (`thresh_pix` = 4), and surrounded by contiguous pixels with values all greater than 3 times the rms noise (`thresh_isl` = 3). These thresholds are obtained from a series of tests by introducing mock sources into the ALMA images and recovering them with PyBDSF. The best performance was evaluated based on completeness and contamination (see Sects. 2.8 and 3.2). Next, PyBDSF fits multiple two-dimensional Gaussians to each "island" depending on the number of peaks identified within it. Thirdly, all Gaussians of the same "island" are grouped into one source, with the summed flux being the integrated source flux, the flux-weighted averaged position being the source position. The total intrinsic source size is obtained via a moment analysis [6] on each individual Gaussian component's intrinsic size [7]. Finally, the errors of each fitted parameters (peak flux, total flux, and each size parameter) are computed using the formulae of Gaussian fitting errors calibrated by Condon (1997).

About 6% of our "islands" are fitted with multiple Gaussians, while the vast majority (94%) has a single Gaussian component. Most of these multi-Gaussian sources are isolated sources but exhibit non-smooth morphologies, either due to resolved spatial components and/or noise in the image. In a few cases, these multi-Gaussian sources might indeed be interacting galaxies. Utilizing information from the prior source catalog, we are able to reliably flag these sources in a later step of our analysis. Therefore they are kept as a single source in the blind source catalog.

Our final blind source catalog is obtained by correcting for flux bias and re-estimating flux errors (see Sect. 3), then applying a primary beam attenuation correction to the photometry of each source (i.e., to the peak flux, total flux, and associated uncertainties). Given that each source represents a high-redshift galaxy with a typical size of 0.5–2″, much smaller than the primary beam, using a single primary beam attenuation correction factor at the source's central position is reasonable.

2.3. *Prior Source Master Catalog*

In addition to the blind source extraction, we utilize known source positions as a prior for the source fitting. This technique allows for deeper detection limits and lowers the spurious source fraction. Before starting the prior fitting, we compiled a "COSMOS master catalog" from a number of multi-wavelength catalogs for sources in the COSMOS field as listed in Table 2. The aim is to be as complete as possible in prior sources while ensuring that source duplication is solved among the various catalogs. Thus we loop over the prior catalogs in the order listed in Table 2. Their respective areal coverage is indicated in Fig. 6. To ensure that a given galaxy (which might be detected in multiple prior catalogs) has only one unique entry in the master catalog, we find out each uniquely-matched group among the prior catalogs (with matching radius 1″) and add into the master catalog only the source coming from the highest-quality (empirically sorted by angular resolution and relative depth) catalog, i.e. listed closest to the top in Table 2.

Our 1″ matching radius corresponds to a worst false-match probability of 13.3% for other catalogs cross-matched to the Laigle et al. (2016) catalog based on Eq. 1 of Pope et al. (2006) [8]. We emphasize that the

[6] See details in http://www.astron.nl/citt/pybdsm/process_image.html.

[7] Each Gaussian component's intrinsic size is their fitted Gaussian size deconvolved with the clean beam which is a two-dimensional Gaussian, and the deconvolution follows the Astronomical Image Processing System (AIPS; Greisen 2003) `DECONV.FOR` module (see also Spreeuw 2010, Chapter 2).

[8] As an additional experiment, we estimated the false-match probability to be 9.9% by first flipping the catalog to be cross-

false-match rate does not affect our photometric work because if a galaxy from another catalog in Table 2 is falsely matched to the Laigle et al. (2016) catalog, we just use the prior position in the Laigle et al. (2016) catalog for our prior photometry. The source position is not forced to be at the exact prior position, as our photometry code will find the best fitting for position and flux (see next section). It may affect our galaxy property analysis via SED fitting in a later step because we use the redshift information from the literature as the prior. However, the influence is minimized by (a) collecting all possible prior redshift information in the literature, (b) verifying via photo-z SED fitting (see Sect. 4.4), and (c) in later steps, we only consider a source a robust galaxy if it passes all our quality assessments (Sect. 4). A falsely-matched source with a wrong prior redshift is unlikely to pass them as detailed in Sect. 4. Yet we can not totally avoid false matches (which should be only a few out of a thousand in our final products), especially when the astrometry in optical/near-IR image data also affects our work (see next section and Appx. A).

The combination of these prior catalogs results in the "COSMOS master catalog" with unique source IDs. In our current master catalog (version `20170426`), because the COSMOS2015 catalog is our primary catalog, all 1,182,108 COSMOS2015 sources are in our "COSMOS master catalog" with the same IDs. 443,688 (37.5%) of them have counterparts in other catalogs. The remaining five catalogs contribute 110,768 new sources which are not in the COSMOS2015 catalog. The Muzzin et al. (2013) catalog contributes 18,536 sources, a fraction of which are from the COSMOS2015 masked regions close to bright stars. The K_s data used for the Muzzin et al. (2013) catalog is shallower than that of the COSMOS2015 catalog (UltraVISTA DR2), so the reason for some new sources should be the different source extraction methods used: the COSMOS2015 catalog uses a z, Y, J, H, K_s-combined χ^2 detection image while the Muzzin et al. (2013) catalog directly uses the K_s image and therefore favors redder sources. The i-band selected catalog contributes 31,159 sources, probably benefiting from its higher angular resolution detection image (see discussion in Sect. 4 of Capak et al. 2007). The IRAC catalog contributes another 4,685 sources. The radio catalog contributes 1,042 sources. Finally, the Sanders et al. (2007) catalog contributes 55,346 sources, but most of them are in an area outside the COSMOS2015 coverage (e.g., Fig. 6), while only 8,893 sources are new in the area covered by both catalogs.

matched to the Laigle et al. (2016) catalog in R.A. positions, then do the cross-match.

The total number of unique priors that fall in primary beam attenuation > 0.2 areas of our dataset version `20180102` (`20180801`) is 41,161 (73,387).

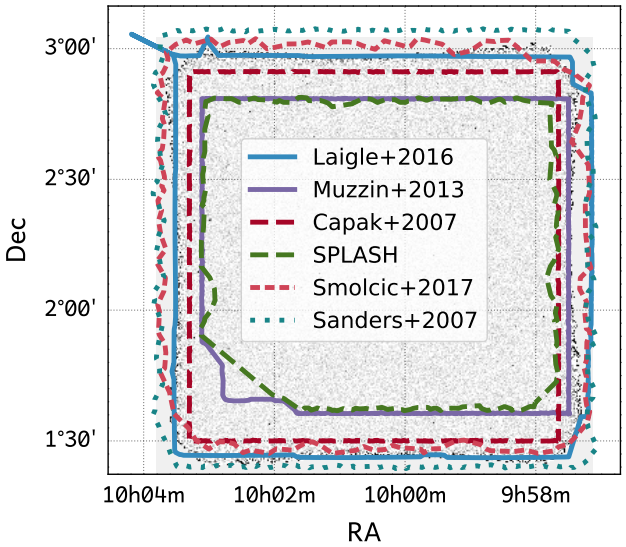

Fig. 6. Coverage of our prior catalogs for the COSMOS field as listed in Table 2. Background image is the *Herschel* PACS 100 μm image, same as in Fig. 2. The colored lines encompass the area searched for sources in the respective prior catalog (see inset for catalog information).

2.4. *Prior Source Fitting*

Utilizing source positions from the COSMOS master catalog, we obtain the (sub-)mm photometry via prior source fitting of the ALMA continuum images. We implement two steps below to optimize the robustness of the fitting. Potential small astrometric inconsistencies between the prior source positions and the ALMA data are taken into account as follows before the full prior source fitting procedure is applied: we calculate the offsets between pre-run ALMA positions and the prior source positions directly from (Laigle et al. 2016) and other prior catalogs, then derive a mean offset for each prior catalog and update all prior positions in our master catalog. Details of the astrometry analysis are given in Appx. A.

As a first step, we identify potential candidate sources based on the S/N of their peak (sub-)mm flux density ($S_{\rm peak}$) or integrated flux density ($S_{\rm total}$). Following Scoville et al. (2014, 2016, 2017), we measure both flux densities in a series of apertures with radii from 0.″25 to 2″ in steps of 0.″25. We follow exactly the Scoville et al. (2016, 2017) method so as to allow for a direct comparison. Using the pixel rms noise calculated from Gaussian fitting to the pixel value distribution of each image, we obtain the S/N ratio for the peak flux density

Table 2. Prior catalogs used for constructing the COSMOS master catalog

Catalog Name (and Reference)	Area ($\deg^2$)	$N_{\rm Catalog}$ [a]	$N_{\rm Master}$ [b]	$N_{\rm Unique}$ [c]	Detection Map	Depth [d]	Res. [e]
COSMOS2015 catalog (Laigle et al. 2016)	1.8	1182108	1182108	738420	VISTA $zYJHK_s$	24.0 ($3\,\sigma$, $3''$, K_s)	$\sim 1''$
K_s-band catalog (Muzzin et al. 2013) [f]	1.6 [f]	263229	18536	10799	VISTA K_s	24.35 ($3\,\sigma$, $2''.1$)	$\sim 1''$
i-band catalog (Capak et al. 2007)	1.8	386125	31159	30146	CFHT $i^\star$ + *Subaru* i^+	26.2 ($5\,\sigma$, $3''$)	$\sim 0''.5$
SPLASH IRAC supplementary catalog [g]	1.6	5390	4685	3690	*Spitzer*/IRAC $3.6+4.5\,\mu$m	25.5 ($3\,\sigma$)	$\sim 1''.6$
VLA catalog (Smolčić et al. 2017)	2	10922	1042	644	VLA 3 GHz	$2.3\,\mu$Jy ($1\,\sigma$)	$\sim 0''.75$
IRAC catalog (Sanders et al. 2007)	2.3	347332	55346	54642	*Spitzer*/IRAC $3.6-8.0\,\mu$m	24.01 ($5\,\sigma$, $3.6\,\mu$m)	$\sim 1''.6$

[a] Total number of sources in each prior catalog. [b] The number of sources in each prior catalog that are not in higher order catalogs (the order is as listed from top to bottom). Which is, the number of sources in our "COSMOS master catalog" that are originated from the current prior catalog. [c] The number of unique sources in each prior catalog, which means these sources have no counterpart in any other prior catalog. [d] The depth is in AB magnitude, and is in an aperture if indicated in brackets. [e] Spatial resolution, or point spread function size of the detection map. [f] The Muzzin et al. (2013) catalog contains sources in the masked area of the COSMOS2015 catalog which are close to bright, saturated stars. [g] Based on the source extraction in the Spitzer Large Area Survey with Hyper-Suprime-Cam (SPLASH; PI: P. Capak) survey data after fitting and removing all COSMOS2015 catalog sources (I. Davidzon; priv. comm.) * The master catalog used in this paper's work has a version code of 20170426.

via $\rm S/N_{peak} \equiv S_{peak}/(\rm rms\ noise)$, and the one for integrated flux density $\rm S/N_{total}$ by dividing $S_{\rm total}$ by the integrated noise in each aperture (i.e., rms noise times the square root of pixel number in each aperture). We refer to this aperture photometry as the GETPIX method hereafter (and compare its results with those from our other photometry methods in Sect. 2.7).

This first GETPIX step also provides guidance for the prior source fitting using GALFIT (Peng et al. 2002, 2010) in the next step. We select $\rm S/N_{total} > 2$ or $\rm S/N_{peak} > 3.6$ sources (same as Scoville et al. 2017) as valid detections. This pre-selection of prior sources is important for applying GALFIT, as it significantly reduces the required computational time for GALFIT by avoiding the fitting of sources that mostly correspond to noise in the image. We have confirmed that this approach is sensible with our Monte Carlo simulations (see Sect. 3). Also, our final catalogs are not sensitive to small changes of these thresholds, because in the end we apply a relatively high $\rm S/N_{peak}$ cut according to our Monte Carlo simulation statistics. Note that in most ALMA images our priors do not have blending issues.

To optimize the GALFIT fitting for source fluxes as well as sizes, an iterative approach is adopted: After the first-pass fitting with point source models to all GALFIT priors fixed at their original positions, we select sources with a fitted magnitude error of < 0.25 [9] or $S_{\rm total} > 3\,\sigma$ (σ being the pixel rms noise) and allow their positions to vary by at most $0''.7$ [10] in the second-pass fitting. Then, in order to identify possible extended sources, we allow sources with fitted magnitude error < 0.20 or $S_{\rm total}$ above 3 times the rms noise to be fitted with circular Gaussian models (and in a next step Sérsic profiles) in the third-pass fitting. We note that our thresholds are very loose and 98% of the sources in our final prior photometry catalog (with a relatively high selection threshold, $\rm S/N_{peak} \gtrsim 5$, according to our Monte Carlo simulation statistics, see Sect. 4.1) are fitted with extended shapes.

For each fit, we ensure that the image background is zero (as already verified by the close to zero means of the distributions of the pixel values from the ALMA images). If a given GALFIT iteration yields bad fits and/or non-convergence, the fitting is repeated with a higher limit for GALFIT iteration [11].

[9] GALFIT fits magnitude instead of flux density

[10] This is the $1\,\sigma$ scatter of the spatial separations between our ALMA sources and their optical/near-infrared counterparts as we examined in Sect. 4.2.

[11] By default, GALFIT iterates a maximum for a total of 100 times, and 10 times when converging to a local minimum. These numbers can be increased to, for example, 1000 total iterations and 255 iterations during convergence, e.g., Liu et al. (2018).

As the GALFIT errors only consider the covariance matrix of the fitting, they do not reflect observational noise or correlated noise. Therefore, we estimate the error in $S_{\rm total}$ ($\sigma_{S_{\rm total}}$) for Gaussian-fitted sources following Condon (1997). This error estimation determines $\sigma_{S_{\rm total}}$ purely from the rms noise, beam major and minor axes FWHM sizes ($\theta_{\rm bmaj.}$ and $\theta_{\rm bmin.}$ respectively), source major and minor axes FWHM sizes ($\theta_{\rm maj.}$ and $\theta_{\rm min.}$ respectively; fitted values and convolved with the beam) and source $S_{\rm peak}$ and $S_{\rm total}$. We further verify that this is in general consistent with our own Monte Carlo simulations (see Sect. 3).

2.5. *Comparing Blind Extraction and Prior Fitting Results*

As a quality check to both blind source extraction and prior source fitting, and to identify potential problem cases, we compare the total fluxes from PyBDSF to those from GALFIT for common sources (within 1.″0 and using the same ALMA images) in Fig. 7. 96% of sources have fluxes agreeing within $3\,\sigma$. Outliers with flux differences of $> 5\,\sigma$ are labeled in the figure. Their PyBDSF and GALFIT fitting models and residuals are further shown and discussed in Appx. B. The three outliers with a GALFIT flux much larger than the PyBDSF flux are caused by poor fits of PyBDSF to their irregular morphologies. The one outlier with a much larger PyBDSF flux than the GALFIT flux is due to a blending of prior sources and given the complex morphology both GALFIT and PyBDSF could not provide an ideal fit.

With both PyBDSF and prior-based GALFIT photometry, we not only obtain accurate independent fluxes which agree very well but also identify those few (0.5% [12]) sources which suffer from source multiplicity/blending issues. These sources need careful visual inspections as well as multi-wavelength diagnostics (e.g., SEDs) in order to fully deblend their ALMA flux, and thus will be analyzed in a future work.

In our released two photometry catalogs, we flag sources for which the total fluxes from the two methods disagree by more than a factor of ∼3.12 ($5\,\sigma$, where σ is the scatter between PyBDSF and GALFIT total fluxes, see Fig. 7) with a column `Flag_inconsistent_flux` and exclude them in subsequent steps. In the next sections, we use the prior photometry flux for the SED fitting. But measurements from both photometry methods will be made public together with the final galaxy SED and property catalog (see Sect. 5).

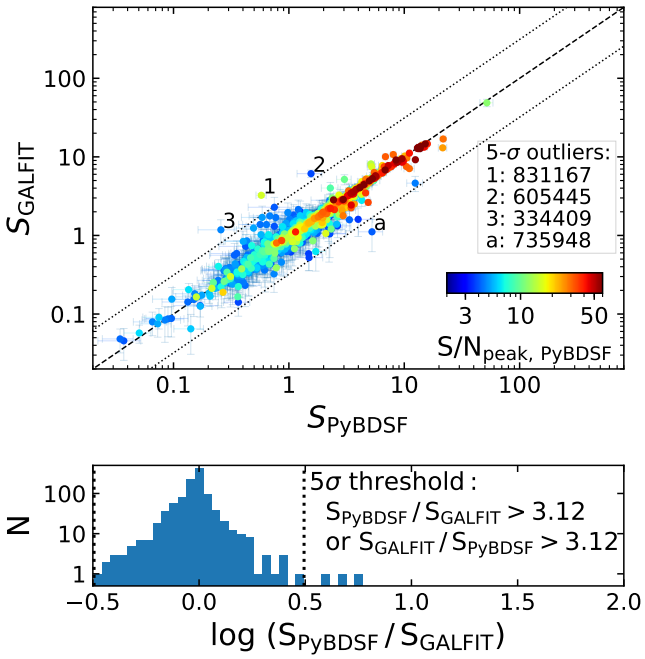

Fig. 7. Top: Comparison of total fluxes derived from the PyBDSF blind source extraction and the prior-based GALFIT source fitting. Data points show sources matched within 1″ and measured on the same image. Color indicates their $\rm S/N_{peak}$, i.e., the ratio of source peak flux to rms noise of the image. The solid line shows the one-to-one correspondence and the two dashed lines indicate the $5\,\sigma$ range, where σ is the scatter measured from the bottom panel. Outliers above $5\,\sigma$ are labeled and discussed in detail in Appx. B. **Bottom:** Histogram of the flux difference on a logarithmic scale, $\log_{10}(S_{\rm PyBDSF}/S_{\rm GALFIT})$. The mean is -0.013 with a standard deviation of 0.10 dex. Dashed vertical lines indicate the same $5\,\sigma$ range as in the upper panel.

2.6. *Comparison to uv-Plane Source Fitting Results*

Instead of measuring the source flux density in the image plane, it can also be directly measured in the *uv*-plane by fitting source models to the visibilities. We use the GILDAS [13] `uv_fit` task to fit Gaussian and/or point source models then compare the total flux with those measured from the image-plane GALFIT and PyBDSF fitting. We verified that GILDAS `uv_fit` gives similar results to the CASA `uvmodelfit` task for high S/N sources (e.g., total flux S/N > 10).

[12] We have 0.5% such sources in our final photometry catalogs selected according to the threshold in Sect. 4.1. This fraction goes up to only 2% if we apply a threshold of $\rm S/N_{peak} \geq 3.0$ to both catalogs.

[13] GILDAS is an interferometry data reduction & analysis software developed by Institut de Radioastronomie Millimétrique (IRAM) and is available from http://www.iram.fr/IRAMFR/GILDAS/. The conversion of ALMA measurement sets to GILDAS/MAPPING *uv* table data follows https://www.iram.fr/IRAMFR/ARC/documents/filler/casa-gildas.pdf.

We run GILDAS `uv_fit` in an iterative approach: first we fit point source models, next for high S/N sources we fit extended Gaussian source models. We fit only for one source at the phase center and allow its position to vary freely by `uv_fit`. In total we ran the *uv*-fitting for 301 pointings from four representative ALMA projects: 2015.1.00137.S, 2013.1.00151.S, 2015.1.00379.S and 2016.1.01208.S (these projects target the dust continuum for hundreds of galaxies from redshift 1 to 3; the PI of the first project is N. Scoville, and for the other three is E. Schinnerer). The `uv_fit` flux densities and the prior-based and blind (sub-)mm photometry agree very well. The difference between blind photometry and `uv_fit` flux densities (on a logarithmic scale) has a median of 0.015 dex and scatter of 0.08 dex. The difference between prior photometry and `uv_fit` flux densities has a median of -0.005 dex and scatter of 0.13 dex, showing a few more outliers (caused by blended priors, same as in Fig. 7).

2.7. *Comparison to Aperture Photometry Results*

We further compare the fluxes from our PyBDSF and prior-based GALFIT fitting with those derived from aperture photometry (Scoville et al. 2016, 2017) (i.e. the GETPIX method described in Sect. 2.4). For sources with GALFIT $S/N_{peak} \geq 3$, the GETPIX method provides flux densities consistent with the ones from GALFIT (the mean of GETPIX to GALFIT flux ratio on a logarithmic scale is 0.004 dex and the scatter is 0.18 dex). Sources with $S/N_{peak} > 10$ are on average biased toward higher GETPIX flux densities, but no more than 10% (the mean value increases to 0.03 dex and scatter 0.06 dex; likely due to bright outlier sources which have non-Gaussian shapes).

The comparison between GETPIX and PyBDSF flux densities yields similar results: for PyBDSF $S/N_{peak} \geq 3$ sources the mean of $\log_{10} S_{\rm GETPIX}/S_{\rm PyBDSF}$ is 0.003 dex with a scatter of 0.15 dex; when considering only $S/N_{peak} \geq 10$ sources the mean is still <0.01 dex. For about 10 sources, we directly compared our flux densities to measurements from Scoville et al. (2016, 2017; priv. comm.), finding similar results to those mentioned above.

2.8. *Inverted Image Fitting and the Fraction of Spurious Detection*

We run our photometry tools (based on PyBDSF and GALFIT) on the inverted images (i.e., the sign of each pixel value is inverted) to estimate the fraction (and probability) of spurious detections by comparing the number of sources detected in inverted images to that in original images. We define the spurious fraction as

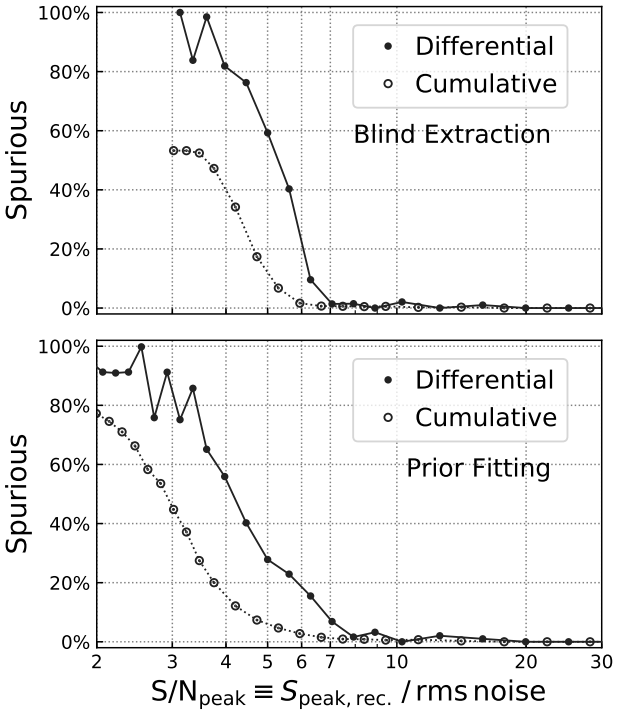

Fig. 8. Fraction of spurious detection for the PyBDSF-based blind source extraction (upper panel) and GALFIT-based prior source fitting (lower panel). The solid curves and filled data points represent the differential spurious fraction at each S/N_{peak} bin, while the dotted curves and open data points represent the cumulative values, i.e., for $S/N_{peak} \geq$ the current bin's S/N_{peak}.

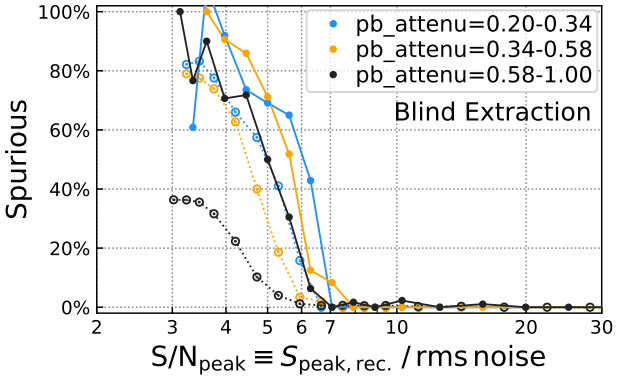

Fig. 9. (This figure has been added.) Spurious fraction for three bins of `pb_attenu`: (0.2–0.34], (0.34–0.58] and (0.58–1.0], which are equally distributed in logarithm. The solid and dashed lines represent differential and cumulative curves respectively (see Fig. 8 caption). We caution that the trend seen here suffers from a strong bias in statistics, because the lowest `pb_attenu` (furthest away from phase center) bin has only about 62 $S/N_{peak} > 5$ detections in original images and 27 in inverted images, while the numbers are ten times larger in the inner most bin with `pb_attenu` $\sim 0.58-1.0$ (although they are equal in area).

the number of sources detected in inverted images compared to the corresponding number in the original images as a function of S/N_{peak} (defined as S_{peak}/rms noise in Sect. 2.4), since this quantity does not depend on any fitted source size.

Since prior fitting needs a prior catalog to proceed with, and because our prior catalog has a very high number density ($\sim$ 700 per arcmin^2) which acts like a random sampling in the image, we directly use our COSMOS master catalog as the prior catalog for the inverted-image GALFIT photometry. The procedure is the same as described in Sect. 2.4, we first run the GETPIX step then iteratively run GALFIT source fitting. In addition, we checked that the spurious detection curve remains the same when shifting the positions of the entire prior catalog by $+/-2''$ in Right Ascension and/or Declination to avoid overlap with real galaxies.

Fig. 8 shows the derived spurious fraction curves as a function of S/N_{peak} for both PyBDSF (top) and GALFIT (bottom) photometry. The differential curve (solid line) indicates the spurious fraction at each S/N_{peak}. The cumulative curve (dotted line) provides the spurious fraction summed over all bins with S/N_{peak} greater than or equal to the current bin. As expected, spurious fractions are lower for the prior-based photometry compared to the blind source extraction due to the availability of information on the presence of a galaxy. Thus the prior-based photometry achieves deeper detection limits.

To investigate whether the primary beam attenuation is affecting the false-positive detection, we have done two tests: one is dividing the spurious fraction curve in bins of primary beam attenuation (`pb_attenu`) as shown in Fig. 9, the other is plotting the radial distribution of all spurious detections from the inverted images in Fig. 10. In the former test, we choose only three bins because of the low number of sources away from the phase center (low `pb_attenu`). We bin in equal ln(`pb_attenu`) intervals which correspond to the same sky area, because `pb_attenu` $\propto \exp(\texttt{dist.}^2/\texttt{pb}^2)$, where `dist.` is the distance of the source to the phase center and `pb` is the FWHM of the primary beam. The spurious fraction decreases when `pb_attenu` becomes closer to 1.0, which is as expected. But we also caution that there is a strong bias in the statistics because the number of sources dramatically differ (see Fig. 9 caption).

In Fig. 10, we show the radial distribution of all sources detected in the inverted images with GALFIT $S/N_{peak} > 2.5$ or found by PyBDSF. Since the spurious fraction curve is slightly higher at larger radii, we might expect the spurious source density to be higher, however, the distribution remains fairly constant out to a `pb_attenu` of $\sim$ 0.3. We attribute the slight drop below $\sim$ 0.3 to the fact that instrumental systematics are likely becoming more prominent, namely (a) the approximation of the primary beam by a Gaussian might no longer be correct [14], and (b) the frequency dependence of the primary beam across the frequency range sampled by the continuum (i.e., 16 GHz between upper and lower boundary of the spectral sidebands) will be more evident at large distances from the phase center. A more detailed investigation is beyond the scope of this paper.

In this work, we provide a photometry catalog out to a primary beam attenuation of 0.2 (i.e. covering the full area of the images that are made available) and provide the `pb_attenu` for each source in our catalog. Note that 91% of our final selected sources lie within a primary beam attenuation of 0.5 and only 2% beyond 0.3. Special care should be applied, e.g., considering a higher S/N_{peak} threshold as shown in Fig. 9 when studying sources below a `pb_attenu` of $\sim$0.5.

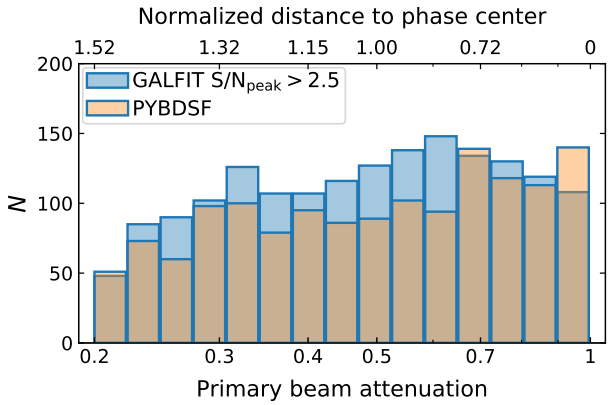

Fig. 10. (This figure has been revised to keep only the histogram.) The radial distribution of spurious detections for the PyBDSF-based blind source extraction (orange) and GALFIT-based prior source fitting (blue). The bottom x-axis is the primary beam attenuation, `pb_attenu`, and the top axis is the normalized distance to the phase center, `dist.`$/(0.5 \times$ `pb`$)$, where `dist.` is the spatial distance to the phase center and `pb` is the FWHM of the primary beam. The histogram bins are equally distributed in ln(`pb_attenu`) (so that the area of each bin normalized by the primary beam area are equal).

3. MONTE CARLO SIMULATIONS

[14] E.g., see https://help.almascience.org/index.php?/Knowledgebase/Article/View/234.

We run extensive Monte Carlo (MC) simulations to verify our two main photometry methods: PyBDSF and GALFIT. The principle idea is to simulate model galaxies and recover them with the same analysis used to create our catalogs. The aims are (1) to test whether the recovered flux densities have a systematic offset to the simulated flux densities, which is hereafter referred to as "flux bias", and to understand its source and quantify it if it exists; (2) to quantify the overall uncertainty on the extracted flux densities, and verify whether the aforementioned Condon (1997) error estimates can statistically describe the uncertainty; (3) to quantify the fraction of sources being recovered from all sources simulated, which is hereafter referred to as "completeness"; and (4) to verify whether the prior information used in the simulations will alter the output statistics or not.

In our simulations, we create artificial sources (of Gaussian shape), insert them into residual images (after blind extraction photometry) and recover them with our photometry pipelines. These steps are repeated several tens to hundreds of times for a large number of images with different properties (details are given in Appx. C). Our artificial sources are created within a grid of input values of both flux density and size. We create two sets of simulations with quite different input distributions defining this grid: (1) We start with a full-parameter-space simulation (hereafter "FULL" simulation) in which the full parameter space of flux density and size is uniformly sampled: S/N_{peak} ranges from 2.5 to 100 in logarithmic intervals, and source major axis size to beam major axis size ratio ranges from 0.1 to 6. Each grid point with a given flux density and size contains the same number of simulated sources. (2) We create another physically-motivated MC simulation, hereafter "PHYS" simulation, where we simulate sources mimicking observed galaxy stellar mass functions (SMFs; e.g., Davidzon et al. 2017), star-forming main sequence relation (MS; e.g., Sargent et al. 2014) and starburst/MS classification (i.e., following the 2-star-formation model (2SFM) of Sargent et al. 2012, 2014; Béthermin et al. 2012a), as well as galaxies' size evolution (e.g., van der Wel et al. 2014; Fujimoto et al. 2017). The motivation for performing our "PHYS" simulation is that galaxies have non-uniform luminosity functions (or number counts) and size distributions. Fainter galaxies are much more numerous than brighter ones, and lower redshift galaxies are in general larger than higher redshift ones. Our comparison of the "FULL" and "PHYS" simulations tests whether the input distribution of the simulations influences the derived recovery statistics.

Due to the large number (1500+) of individual ALMA imaging data, we select a subset (150+) of representative images for each Scheduling Block of each Science Goal in each ALMA project. In this way we make sure that all different observing scenarios (frequencies, spatial resolutions, integration times, etc.) are covered.

For each selected image, we perform the "FULL" and "PHYS" type simulations 4225 and 273 times respectively, depending on the grid of simulation (see Appx. C), resulting in 4225 and ∼3000–25000 simulated objects respectively. The number of sources in the "PHYS" simulation varies with the image field of view and the observing wavelength, and dominates with fainter sources due to the assumed galaxy stellar mass functions and main-sequence correlation as well as the SEDs. Details of the two simulations are presented in Appx. C.

We then recover the simulated objects with our PyBDSF and GALFIT photometry pipelines respectively, using the identical settings as for the real ALMA data. Therefore, we have four sets of simulated-and-recovered data to analyze and compare: FULL-PyBDSF, FULL-GALFIT, PHYS-PyBDSF and PHYS-GALFIT.

In the next sections, we discuss the flux bias and flux errors for each simulation set, and characterize them by two normalized parameters: the fitted source peak flux density normalized by the rms noise,

$$S/N_{peak} \equiv S_{peak}/\text{rms noise} \quad (1)$$

and the fitted source area (convolved with the beam) normalized by the beam area,

$$\Theta_{beam} \equiv \sqrt{\text{Area}_{source,\,convol.}/\text{Area}_{beam}} \quad (2)$$

Note that the different type of simulations yield clear differences in the parameters of interest, especially the flux bias correction, as we will show in the following when comparing the results from all four simulated data sets.

3.1. *Analyses of the "FULL" and "PHYS" Simulations*

Although the simulated total source flux density, $S_{sim.}$, overall agrees well with the recovered total source flux density, $S_{rec.}$, a substantial bias between $S_{sim.}$ and $S_{rec.}$ becomes obvious when looking at the dependency on the flux S/N. When normalizing the difference between $S_{sim.}$ and $S_{rec.}$ by the measured flux error, the histogram distribution of $(S_{sim.} - S_{rec.})/\sigma_{S_{rec.}}$ exhibits a non-zero mean and non-unity scatter (such histograms are illustrated later in Appx. C.1.2, C.2.3 and C.3). This indicates that the measured fluxes need to be corrected

for flux biases, and the errors in the measured fluxes need to be re-estimated.

To analyze the flux bias and errors from our simulations, we bin all simulated and recovered sources in the 2D parameter space of S/N_{peak} and Θ_{beam}, and consider flux bias and error to be functions of these two parameters (Condon 1997; Bondi et al. 2003, 2008; Schinnerer et al. 2010; Jiménez-Andrade et al. 2019). Because S/N_{peak} and Θ_{beam} are both normalized quantities, sources from different ALMA projects can be combined.

For each S/N_{peak} and Θ_{beam} bin, we compute the mean and median of the relative flux density difference $((S_{sim.} - S_{rec.})/S_{sim.})$. The flux bias is then defined as:

$$\eta_{bias} \equiv \langle (S_{sim.} - S_{rec.})/S_{sim.} \rangle \quad (3)$$

which represents how the recovered flux density is biased relative to the simulated flux density. The corrected flux density can then be calculated as:

$$S^{corr.}_{rec.} = S_{rec.}/(1.0 - \eta_{bias}) \quad (4)$$

We note that computing the flux bias using the noise-normalized flux density difference $((S_{sim.} - S_{rec.})/\text{rms noise})$ leads to no obvious difference.

Then we also compute the scatter of $(S_{sim.} - S_{rec.})/\text{rms noise}$ (we computed the standard deviation and the lower and higher 68^{th} percentiles, see Sect. 3.1.3) and denote it as:

$$\eta_{error} \equiv \sigma_{[(S_{sim.} - S^{corr.}_{rec.})/\text{rms noise}]} \quad (5)$$

We do not use the relative difference $((S_{sim.} - S_{rec.})/S_{sim.})$ because its scatter has an asymmetric distribution. The corrected flux density error can then be computed as:

$$\sigma_{S^{corr.}_{rec.}} = \eta_{error} \times \text{rms noise} \quad (6)$$

Combining all bins, we can measure η_{bias} and η_{error} as functions of S/N_{peak} and Θ_{beam}, which are illustrated in Fig. 11 for "FULL" simulation with PyBDSF recovery as the example (other three simulation-recovery pairs are analyzed similarly and the Θ_{beam}-collapsed figures can be seen in Fig. 12 and 13). This figure demonstrates that the flux bias and error do strongly correlate with S/N_{peak} and Θ_{beam}.

3.1.1. *Flux bias in the PyBDSF photometry*

The PyBDSF photometry measurement of a source always includes the intrinsic source flux plus a contribution from noise, thus it always fits positive source fluxes, and the measured fluxes are statistically boosted by a certain amount which we define as the flux bias.

Based on our simulations, we characterize the flux bias correction factor (η_{bias}, Eq. 3) by the two measurable parameters S/N_{peak} and Θ_{beam}, and apply the flux bias correction to the measured/recovered flux with Eq. 4. We find these two parameters to much more strongly affect the flux bias than other parameters, e.g., absolute source size or beam size. After the flux bias correction, the extracted total fluxes for simulated sources in maps of different spatial resolutions exhibit no obvious further bias from their simulated total fluxes. Here, we also found that the flux bias parametrization strongly depends on the input mock source populations of the MC simulation as demonstrated below.

In Fig. 12, we compare the flux bias of the PyBDSF photometry characterized from our "FULL" and "PHYS" simulations. S/N_{peak} is on the x-axes and Θ_{beam} is indicated by the color. The flux bias is a strong function of both S/N_{peak} and Θ_{beam}. It rapidly becomes significant with decreasing S/N_{peak}. For example, $|S_{sim.} - S_{rec.}|$ can be $> 10\%$ of $S_{sim.}$ when $S/N_{peak} \lesssim 10$. Secondly, sources with larger sizes suffer a stronger flux bias: a source with a measured size 4 times the beam size can be boosted by $\sim$80% of $S_{sim.}$ at a $S/N_{peak} = 5.77$ (where the spurious fraction at this S/N_{peak} is $\sim$40%, see Fig. 8); while a unresolved source is only boosted by $\sim$20% at the same S/N_{peak}.

The flux bias functions derived from the two simulations are fully consistent at the bright end, e.g., $S/N_{peak} \gtrsim 10 - 20$. Discrepancies between the simulations become only obvious at the faint end of S/N_{peak} for sources with small Θ_{beam}. The flux bias in the "FULL" simulation is much smaller than compared to the "PHYS" simulation. This is due to the difference of the input populations of the two simulations. The effect of "resolution bias" is likely the main reason – such a bias causes sources with low S/N and large *simulated* sizes to have much smaller *recovered* sizes or even be unresolved (or undetected) and also causes their fluxes to be underestimated instead of boosted by noise. This is common in radio photometry, where the spatial resolution is comparable to and even smaller than the sizes of galaxies at high redshift, e.g., as discussed in Bondi et al. (2003, 2008). The resolution bias is much more evident at the faint end of the "FULL" simulation than the "PHYS" simulation because of the higher number of large sources simulated in the former case. More discussion is presented in Appx. C.1.3.

In reality, the physical sizes of galaxies increase with cosmic time and scale with stellar masses (van der Wel et al. 2014), and their angular sizes (stellar component) increase quickly from $z \sim 1$ to the present. This means lower redshift galaxies with high stellar masses tend

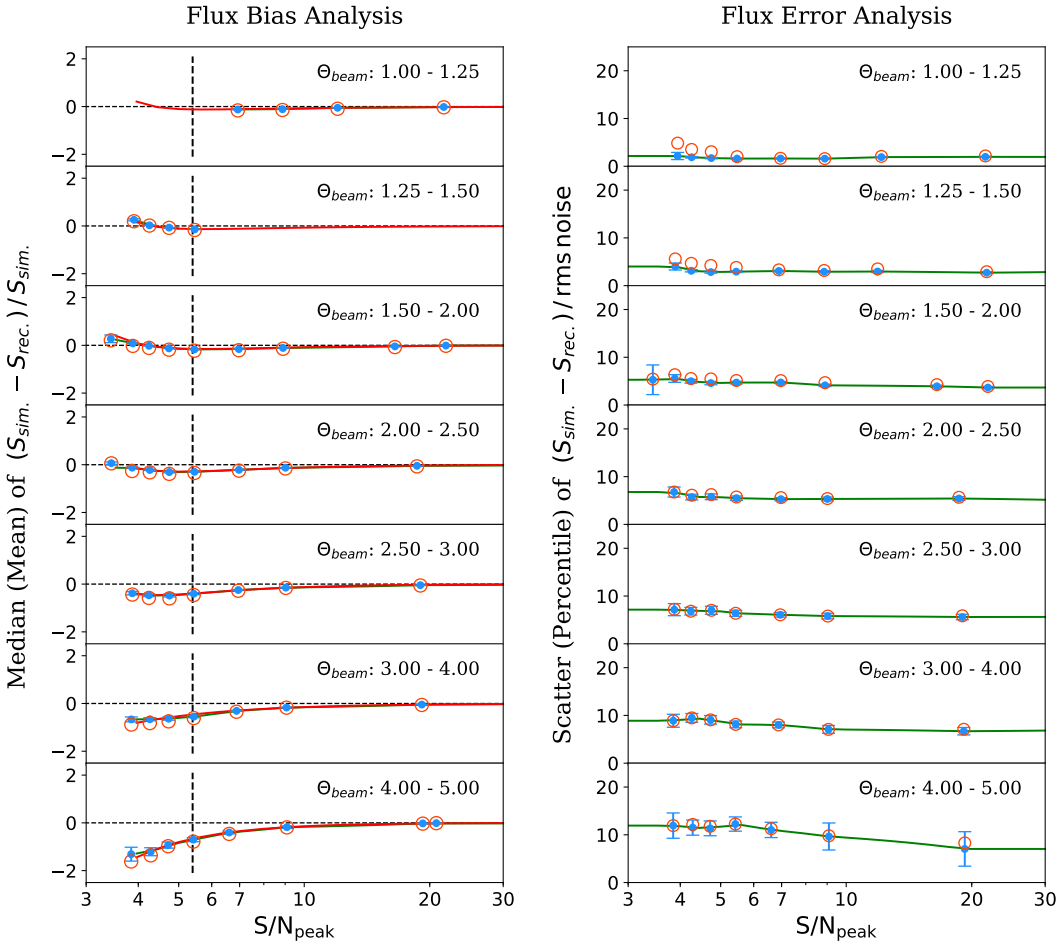

Fig. 11. PyBDSF flux bias and flux error as functions of measured S/N$_{\rm peak}$ (Eq. 1) and $\Theta_{\rm beam}$ (Eq. 2) in the left and right panels, respectively, from the "FULL" simulation. Sub-panels (from top to bottom) are bins with increasing $\Theta_{\rm beam}$ (as labeled). In the left panels, red and blue circles correspond to the mean and median of $(S_{\rm sim.} - S_{\rm rec.})/S_{\rm sim.}$, respectively. Solid red lines are function fitting (with the form $a\,{\rm S/N}_{\rm peak}^{m} + b\,{\rm S/N}_{\rm peak}^{n}$) to the flux bias data points (but no feasible function form could be fitted for flux error), and solid green lines are interpolations or extrapolations (visible in the right panels). The vertical dashed line corresponds to our sample selection threshold as will be detailed in Sect. 4.1. In the right panels, red and blue circles represent the scatter (standard deviation) and (the minimum of upper and lower) 68$^{\rm th}$ percentile of $(S_{\rm sim.} - S_{\rm rec.})$/rms noise, respectively. See text in Sect. 3.1.

to be largest. These galaxies can be bright at radio wavelengths but are in general much fainter and even undetectable at (sub-)mm wavelengths (due to the K-correction and the general drop in star formation activity). Therefore in our ALMA (sub-)mm data, the real galaxy angular size distribution should be dominated by small sources, i.e., it is better described by the "PHYS" simulation rather than the "FULL" simulation. Therefore we use "PHYS" simulation-based flux bias functions for the final correction of the photometry.

3.1.2. *Flux bias in the* GALFIT *photometry*

In Fig. 12, we show the flux bias parametrizations derived for the GALFIT photometry based on both simulations. Similar to the PyBDSF photometry, the GALFIT photometry also shows both flux-boosting due to noise and flux-underestimation due to the "resolution bias".

The GALFIT photometry has a smaller flux bias, which is likely due to the use of known prior position information for the photometry and the optimized iterative photometry approach (Sect. 2.4). It even achieves a better accuracy for sources with largest measured sizes ($\Theta_{\rm beam} \sim 5$) than those with slightly smaller measured sizes ($\Theta_{\rm beam} \sim 3$), if their S/N$_{\rm peak}$ are above 20 or so.

3.1.3. *Flux error estimation for PyBDSF photometry*

With Eq. 5 and 6, we estimate the flux error factor ($\eta_{\rm error}$) from our simulation bins and parameterize it by S/N$_{\rm peak}$ and $\Theta_{\rm beam}$ (after the correction for flux bias). We compute $\eta_{\rm error}$ in a given bin by computing both

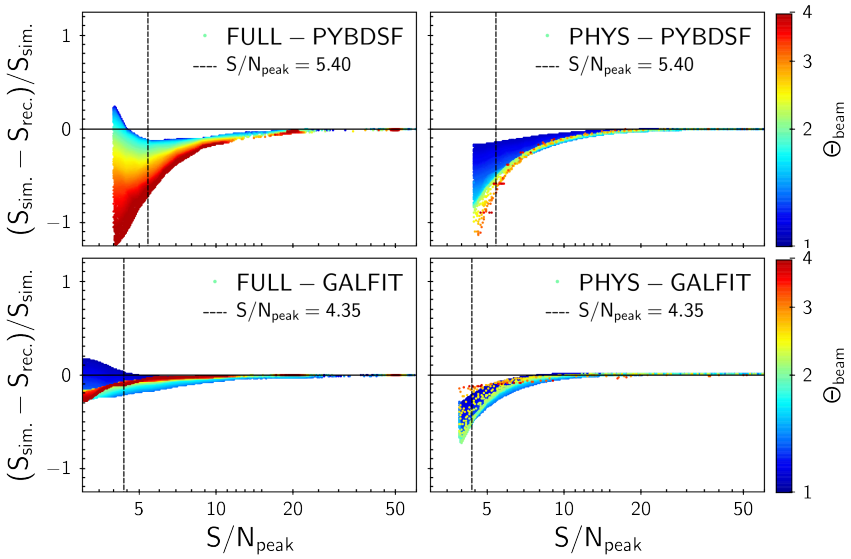

Fig. 12. The flux bias (as defined in Eq. 3) versus the measured S/N$_{peak}$ (as defined in Eq. 1) statistics for PyBDSF (*top panels*) and GALFIT (*bottom panels*) photometry, each based on our two types of Monte-Carlo simulations (*left* panels are "FULL" simulation and *right* panels are "PHYS" simulation; see the label in each panel). Color represents the geometric mean source-to-beam size ratio (Θ$_{beam}$, as defined in Eq. 2 in Sect. 2.2) and is the same in all four panels. Vertical lines are our S/N$_{peak}$ thresholds for the sample selection in Sect. 4.1.

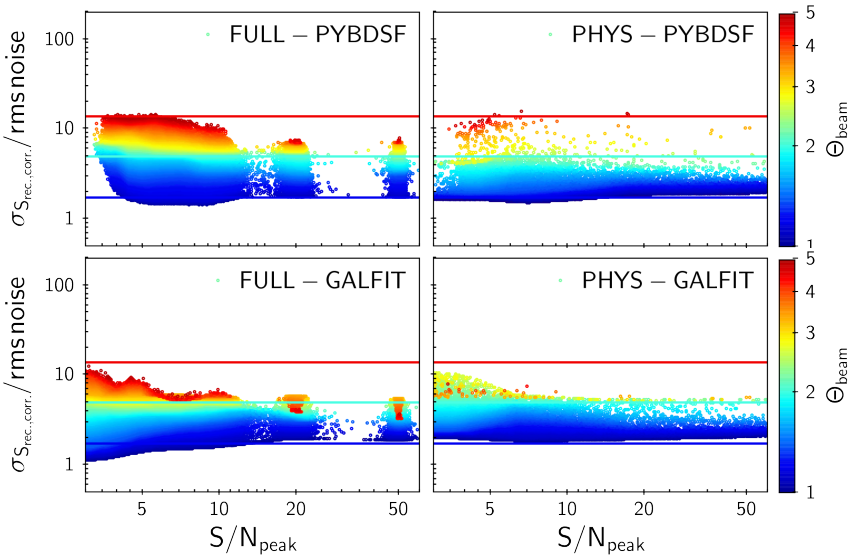

Fig. 13. Similar to Fig. 12 but shows the flux error factor η_{error} (as defined in Eq. 5) versus the measured S/N$_{peak}$ (as defined in Eq. 1 in Sect. 2.2) for our simulated sources. Statistics for the two photometry methods (PyBDSF: *top*; GALFIT: *bottom*) based on our two types of simulation ("FULL": *left*, "PHYS": *right*) are shown. Color represents the source-to-beam area ratio (Θ$_{beam}$, as defined in Eq. 2) and is the same in all panels. The horizontal colored lines show the expected flux errors using the Condon (1997) prescription for Θ$_{beam}$ = 1, 2 and 5 (same color-coding as the data points). See text for further details.

the standard deviation and the upper and lower 68th percentiles. Because the data do not usually follow a normal distribution in ($S_{sim.} - S_{rec.}$), both of these error estimates do not always agree with each other. This can be seen in the right panels of Fig. 11, especially for low S/N$_{peak}$ data points, where the standard deviation is usually larger than the one derived from the percentiles. And we find that the minor value of the upper and lower 68th percentiles can better represent the underlying scatter (which are shown in later figures).

Condon (1997) proposed a mathematical recipe for estimating the errors of a six-parameter Gaussian fit with correlated noise. As shown by their Eqs. 32, 41 and 42, the total flux error can be characterized by the following parameters: the convolved source size parameters (major and minor axes FWHM sizes, denoted as $\theta_{maj.}$

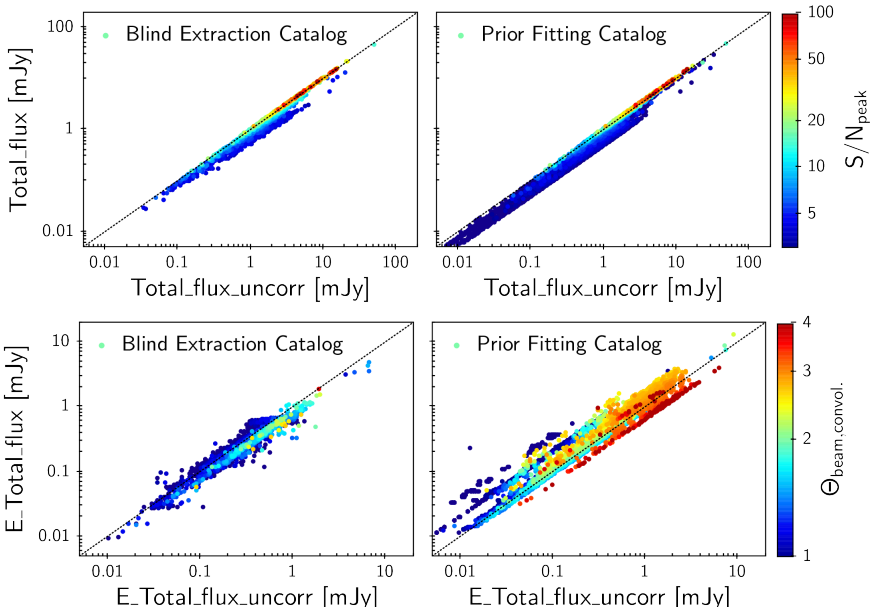

Fig. 14. Comparison of the final corrected and uncorrected fluxes (upper panels) and errors (lower panels) for real data's blind extraction and prior fitting photometry catalogs (left and right panels in each row respectively). The two upper panels have the same axes ranges and color bar indicating the measured S/N_{peak}, and similar for the two lower panels but with the color bar indicating the measured Θ_{beam}.

and $\theta_{min.}$ respectively, corresponding to θ_M and θ_m respectively in Condon 1997), the beam size parameters (major and minor axes FWHM sizes, denoted as $\theta_{bmaj.}$ and $\theta_{bmin.}$ respectively, corresponding to θ_N and θ_n respectively in Condon 1997), and the measured total flux (S_{total}). Such a recipe has later been adopted in Bondi et al. (2003, 2008), Schinnerer et al. (2010) and Smolčić et al. (2017) for the VLA source fitting photometry.

In this work, because our ALMA data have different beam sizes, we express these size quantities in the normalized form: the geometric mean of the source size normalized by the beam size, Θ_{beam} as defined in Eq. 2 which equals $\Theta^2_{beam} \equiv (\theta_{maj.}\theta_{min.})/(\theta_{bmaj.}\theta_{bmin.})$; the size of the source major axis normalized by the beam, $\Theta_{maj.} \equiv \theta_{maj.}/\theta_{bmaj.}$; and the size of the source minor axis normalized by the beam, $\Theta_{min.} \equiv \theta_{min.}/\theta_{bmin.}$.

Because the total flux is the product of peak flux and source area, we can write:

$$S_{total} = S_{peak} \times \frac{(\theta_{maj.}\theta_{min.})}{(\theta_{bmaj.}\theta_{bmin.})} \equiv S_{peak} \times \Theta^2_{beam} \quad (7)$$

Therefore, the (Condon 1997; C97) recipe can be rewritten as:

$$\frac{\sigma^2_{S_{total},\,C97}}{[\text{rms noise}]^2} = \left(\Theta^2_{beam} \times \frac{8}{[1+\frac{1}{\Theta^2_{maj.}}]^{1.5}[1+\frac{1}{\Theta^2_{min.}}]^{1.5}}\right) + \frac{8}{[1+\frac{1}{\Theta^2_{maj.}}]^{2.5}[1+\frac{1}{\Theta^2_{min.}}]^{0.5}} + \frac{8}{[1+\frac{1}{\Theta^2_{maj.}}]^{0.5}[1+\frac{1}{\Theta^2_{min.}}]^{2.5}} \quad (8)$$

Condon (1997) validated the coefficients/indices in their equations using ∼3000 simulations. Because our ALMA photometry is more diverse than their simulations in both data complexity (the variety of beam size, rms noise) and photometry method (e.g., involving iterations), we need to verify that the Condon (1997) recipe is still appropriate for our analysis.

In Fig. 13, we present how our estimated η_{error} changes with S/N_{peak} and Θ_{beam}, and compared with the Condon (1997) errors (horizontal lines). The four panels show the same diagram for our two photometry methods and the two simulations.

According to Eq. 8, the flux error normalized by the rms noise should be independent of S/N_{peak} but strongly dependent on Θ_{beam}. Fig. 13 indeed shows a strong de-

pendency on $\Theta_{\rm beam}$ but also indicates a weak dependency on $S/N_{\rm peak}$. For sources with small sizes (relative to the beam), the flux error becomes larger for larger $S/N_{\rm peak}$ (by about 15% within the range indicated in the figure). However, for sources with large sizes (relative to the beam), it becomes smaller for larger $S/N_{\rm peak}$ (by about 40% within the range of the figure).

The expected Condon (1997) errors for $\Theta_{\rm beam} = 1$, 2 and $\Theta_{\rm beam} = 5$ cases are shown as horizontal lines in Fig. 13, computed using Eq. 8 and assuming a minor/-major axis ratio of 1. Note that a smaller axis ratio will lead to a smaller Condon (1997) error value (by about 15% for $\Theta_{\rm beam} = 5$ when reducing the axis ratio from 1 to 0.1. Our simulation-derived errors (colored data points) are consistent with Condon (1997) errors (colored lines) at the low $S/N_{\rm peak}$ end and at smallest and largest sizes (represented by the colors). However, the "FULL" simulation panel indicates that Condon (1997) errors are overestimated by about 40% for large, high $S/N_{\rm peak}$ sources; while the "PHYS" simulation panel indicates that Condon (1997) errors are underestimated by about 15% for small, high $S/N_{\rm peak}$ sources. Both simulations show that the Condon (1997) errors are slightly overestimated at $S/N_{\rm peak} \sim 5-10$ for small- and intermediate-sized sources.

In our final catalog, we provide both our simulation-derived total flux errors and those given by our photometry pipelines which are based on Condon (1997) [15].

3.1.4. *Flux error estimation for GALFIT photometry*

The flux errors are analyzed in a similar way for GALFIT photometry. The same diagnostic plots are shown in the bottom panels of Fig. 13. The trends for the GALFIT photometry is very similar as for PyBDSF. The GALFIT photometry has even smaller flux errors for large size sources than PyBDSF photometry. Both methods involve multiple iteration or multi-source fitting (rather than one-time simple 2D Gaussian fits), and thus the reason for these trends is not very clear. Yet the different inputs for the two types of simulations do not have a sizeable impact here.

3.1.5. *Final Corrections*

We finally correct flux biases and re-estimate flux errors for both the simulation catalogs and the real data's blind extraction and prior fitting catalogs, based on our aforementioned recipes (as functions of $S/N_{\rm peak}$ and $\Theta_{\rm beam}$; Eqs. 4 and 6, respectively). We choose the "PHYS" simulation for the final correction, considering the discussion in the previous sections, i.e., "PHYS" simulation is more representative of our real data. Note that using "FULL" simulation would underestimate the flux bias correction hence lead to larger fluxes especially for large sources.

The comparison of corrected and uncorrected fluxes and errors for real catalogs are shown in Fig. 14. Based on which, we find that our corrected fluxes and errors follow well-behaved statistics (see details in Appx. C.3), which means flux biases (e.g., flux-boosting) are fully removed and flux errors can fully reflect the scatters of photometry measurements introduced by the noise in the data.

Further in Figs. 15 and 16 we present the distributions of primary-beam-corrected total flux and fitted intrinsic size versus source peak-to-rms noise $S/N_{\rm peak}$ (see Eq. 1), beam-normalized source size $\Theta_{\rm beam}$ (see Eq. 2), and the rms noise and beam major-axis FWHM of the ALMA data. These figures show that our detections span a large range in flux and size. Note that the continuum wavelengths of the ALMA detections also vary: about 44% of the data are at $\sim 870\,\mu$m, about 49% at $\sim 1.0-1.5$ mm (mostly 1.25 mm), $\sim 1\%$ at $\sim 1.9-2.3$ mm and $\sim 6\%$ at $\sim 2.5-3.4$ mm. Thus the sensitivity shown cannot straight-forwardly be compared to single-band ALMA continuum surveys. From these figures, good consistency between the two photometry catalogs is also evident. The prior catalog extends to a slightly fainter regime and only a minor fraction of sources are fitted with smaller sizes. As the aim here is to obtain good continuum photometry catalogs, the study of the uncertainty on source sizes is the topic of future work.

3.2. *Completeness*

In this section, we analyze the completeness of our photometry by examining the fraction of simulated sources that are successfully recovered to the total simulated number. The photometry is incomplete for several reasons: (1) some faint sources are undetected due to noise fluctuation; (2) PyBDSF groups blend multiple sources into one source; (3) GALFIT might give wrong best-fit results in case of severely clustered priors; and (4) PyBDSF has certain flagging criteria to filter out non-physical sources [16]. To assess the contribution of

[15] Note that in PyBDSF, if a source is fitted with a single-Gaussian component, then its total flux error is based on Condon (1997); but if it is fitted with multiple Gaussian components, then the error is propagated.

[16] According to the PyBDSF documentation http://www.astron.nl/citt/pybdsf/process_image.html#flagging-opts,

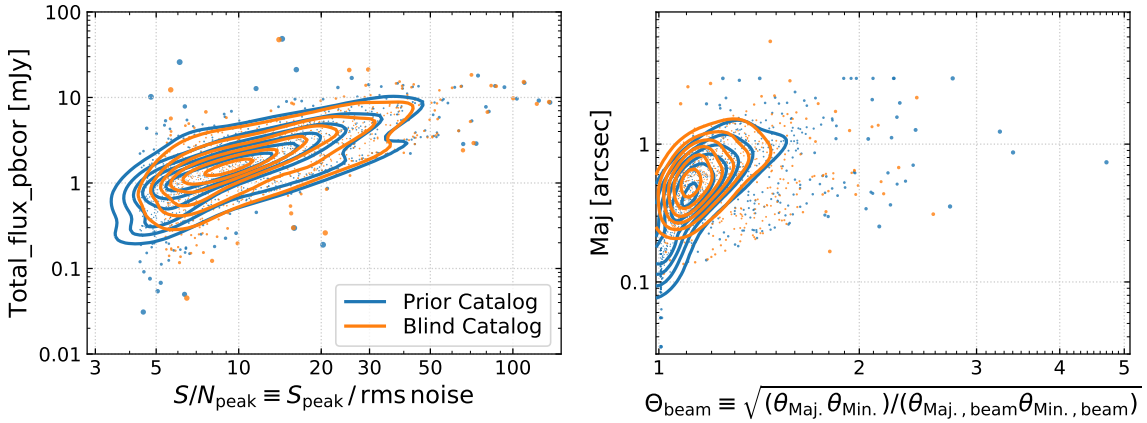

Fig. 15. Total flux (primary beam corrected) versus source peak flux to rms noise ratio S/N_{peak} (see Eq. 1) (*left panel*) and the fitted intrinsic source major-axis FWHM versus the beam-normalized source size Θ_{beam} (see Eq. 2) (*right panel*) for the ALMA detections with $S/N_{peak} > 4.35$ and 5.40 in our two prior- and blind-photometry catalogs respectively (the S/N_{peak} thresholds are determined in Sect. 4.1). Contours are the density of the data points. The size of a data point scales inversely with the density for illustration purpose.

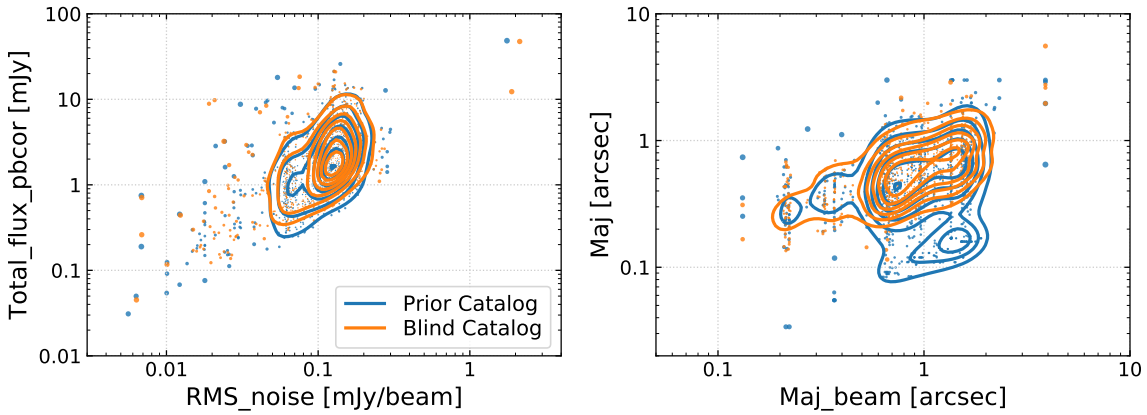

Fig. 16. Total flux (primary beam corrected) versus rms noise (*left panel*) and the fitted intrinsic source major-axis FWHM versus beam major-axis FWHM (*right panel*). Contours and data points are in the same style as in Fig. 15.

these effects, we calculate the completeness curves as a function of S/N and source sizes (normalized by the beams).

We use both PHYS and FULL simulations to verify the completeness. Note that the two simulations have very different source flux and spatial distributions. Sources are isolated and have flat flux distribution in FULL simulation. While in the PHYS simulations, sources have instead realistic spatially distribution as well as a flux distribution that fully agrees with the observed mm number counts (see Appx. C.2).

We cross-match the PyBDSF source recovery catalog to the simulated catalog for each image by coordinate using a search radius of 1.5″ [17], and match by ID for the GALFIT recovery catalog. We measure the completeness as the ratio of the number of sources in the cross-matched catalog and those in the simulated catalog for each bin of S/N_{peak} and Θ_{beam}. We confirmed that the wide range in rms noise and beam size does not affect the completeness estimates by splitting our simulations in random half. Using a smaller search radius has a very minor effect as only 4% (10%) sources have recovered position shifted by more than 1.0″ (0.6″) from the simulated position.

Moreover, the completeness is associated to certain detection criteria. Within PyBDSF, the detection is defined as an extracted source which passes `thresh_pix`, `thresh_isl` and other flagging criteria. Therefore, the

PyBDSF flags apparently non-physical sources. See more details therein.

[17] This corresponds to a false-match probability of $\leq 1.3\%$ for PHYS simulations according to Eq. 1 of Pope et al. (2006).

remaining discussion within this section is focused on the PyBDSF setups (Sect. 2.2). In GALFIT, a detection is slightly more complex to define, as GALFIT always fits a positive flux density for each prior. Thus, we apply a S/N cut to the GALFIT catalog before computing the completeness (without such a S/N cut, the recovery rate would be 100% as every prior is fitted with a flux density).

In the left panels of Fig. 17, we show the completeness curves for the PyBDSF photometry as a function of S/N_{peak}. As sources tend to be small relative to the beam size (with a median (mean) observed size of $\Theta_{beam,\,sim.\,convol.} \sim 1.2$ (1.6)) in the "PHYS" simulation (top-left panel), we do not distinguish between source sizes . The "FULL" simulations (bottom-left panel) have sufficient statistics to study the effect of source sizes, thus we show completeness curves for different simulated source sizes in the bottom-left panel. Here, we consider simulated size instead of recovered size, as the latter is unavailable for undetected sources. Large sources are slightly more complete than small sources at very low $S/N_{peak} \sim 2-4$. This trend reverses at a higher S/N_{peak} (up to $S/N_{peak} \sim 20$) above which the completeness for sources reaches ∼100%. In principle, at a given S/N_{peak}, sources with larger recovered size should have higher completeness. We speculate that the previously discussed resolution bias, spatial noise fluctuation and the "island" feature of PyBDSF all play a role at the low-to-intermediate S/N_{peak} regime — a larger simulated source is easier to detect due to a higher number of pixels above the threshold, but at the same time it has a chance being recovered with a smaller size or even as an (or multiple) unresolved source(s) by PyBDSF (especially for the largest simulated sizes). Thus these effects lead to a lower completeness for the largest simulated sources even at $S/N_{peak} \sim 10-20$. While fine tuning the PyBDSF parameters can achieve better detection for large sources, this would require more dedicated effort beyond our systematic approach which is tailored to the bulk of source properties expected. Moreover, our prior photometry is fitting well large sources ($< 3''$), thus such cases will be identified when we cross-match the prior- and blind-photometry catalog (see Sect. 4.1), and currently no such source is found in our dataset as we excluded beam $< 0.1''$ ALMA data.

The shaded areas in Fig. 17 indicates an uncertainty of a factor of two in the estimated incompleteness in "PHYS"–PyBDSF, and are the same in all other panels. Comparison between the completeness for the smallest sources in the "FULL" simulation and the one from the "PHYS" simulation gives a $\sim 3\%$ lower completeness at $S/N_{peak} \sim 20-40$. This difference is caused by source blending and exactly corresponds to the 3.5% multi-Gaussian sources detected in our data set. As described in Sect. 2.2, when several sources are blended, PyBDSF fits multiple Gaussians and groups them as one island which is then output as a single source.

In the right panels of Fig. 17, we show how different ALMA beam sizes (absolute values in units of $''$) would impact the completeness. We find that as long as the ALMA beam is between $0.2'' - 1''$, the completeness is not obviously affected. For ALMA beams larger than $1''$, completeness drops by $\sim 5-10\%$ even for a high $S/N_{peak} \sim 20-50$ source in our PHYS simulation, and is likely because sources are clustered and large ALMA beam starts to cause blending effect, and also because PyBDSF has the "island"-grouping feature (Sect. 2.2).

In addition, in Appendix Fig. C.4 we show the completeness as 2D functions of both S_{peak}/rms noise and S_{total}/rms noise and Θ_{beam}. We find a good agreement between our completeness analysis and similar work by Jiménez-Andrade et al. (2019) for PyBDSF photometry in their COSMOS VLA data as well as by Franco et al. (2018) for Blobcat (Hales et al. 2012) photometry on their ALMA deep field data. (Further we discuss the comparisons of our completeness to other (sub-)mm/radio photometry works: Karim et al. 2013; Ono et al. 2014; Aravena et al. 2016; Hatsukade et al. 2016; Franco et al. 2018), which confirm that more realistic simulations are required to better recover the statistics).

Given our finding that completeness shows an obvious dependency on source sizes, if selecting a sample with a total flux S/N threshold, the sample will have different completeness for different sizes. But when selecting with a constant S/N_{peak} threshold, the sample will have a homogeneous completeness. Thus we use S/N_{peak} to select our final sample (see the next section). Furthermore, we confirm that the spurious fractions derived from the simulations are consistent with those based on inverted image fitting in Sect. 2.8. The robust estimates of the fractions of completeness and spurious sources provide us with a good handle of the performance of our photometry methods. For a given S/N_{peak}-selection threshold, we know how many real sources are missed and how many could be spurious. While there is no way to improve on the non-detections, there are a number of automated examinations that can significantly reduce the number of spurious sources in our final galaxy catalog (see next sections).

4. GALAXY CATALOG AND PROPERTIES

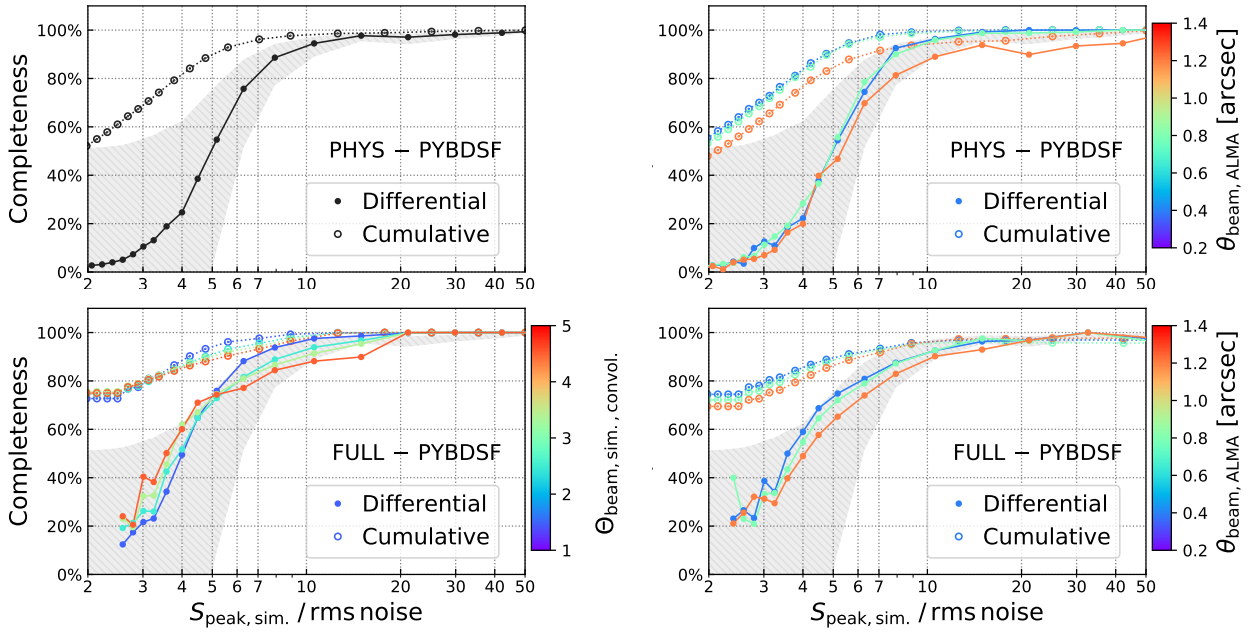

Fig. 17. Completeness of the PyBDSF source extraction as a function of S/N_{peak} based on our simulations ("PHYS": *top panels*; "FULL": *bottom panels*). Color in the *bottom left panel* represents simulated source size (convolved, normalized by the ALMA beam, i.e., Θ_{beam} as defined in Eq. 2) and color in the *top and bottom right panels* are both absolute ALMA beam sizes (θ in units of $''$). Differential completeness at a given S/N_{peak} is marked by solid symbols, while the cumulative completeness for the range above a given S/N_{peak} is shown by open symbols. The shaded areas indicate a factor of two uncertainty in the incompleteness for the "PHYS" simulations in the first panel and is repeated in other three panels for comparison.

In this section, we discuss the selection of reliable ALMA detections from the two photometry catalogs and the construction of our 'galaxy catalog'. Given the extensive information on galaxies in the COSMOS field that is available in the literature, we have devised rigorous inspections to ensure that our galaxy sample and its SEDs are reliable. These inspections include the identification of spurious sources and galaxies with inconsistent photometric and/or spectroscopic redshifts in the literature. We further discuss how galaxy properties are obtained via multiple SED fitting techniques including consistency and reliability checks. The work flow of this analysis step (including Sect. 4.1 to 4.6) is illustrated in Fig. 18.

4.1. *Combining the two photometric catalogs*

We apply a S/N_{peak} cut at 5.40 to our blind source extraction catalog (Sect. 2.2) and a S/N_{peak} cut at 4.35 to our prior source fitting catalog (Sect. 2.4). These thresholds are selected such that the differential spurious fractions are both 50% at the applied S/N_{peak} cut level, and the cumulative spurious fractions are <8% and <12% for the blind- and prior-selected samples, respectively (see Sect. 2.8 and Fig. 8). The corresponding differential completeness at those thresholds are 57% and 98%, and the cumulative ones are as high as >92% and >99%, respectively (see Sect. 3.2 and Figs. 17, C.4). In Fig. 19, we show the S/N_{peak} histograms of the blind and prior catalogs and the applied thresholds.

To merge the two photometric catalogs, we spatially cross-match their sources with a radius of $1.0''$ (false-match probability 0.5% applying Eq. 1 of Pope et al. 2006; see also further discussion of the counterparts association in the next section), and we find 820 sources in common. Another 326 sources are only present in one catalog (207 sources in the prior catalog and 119 sources in the blind catalog). The S/N_{peak} histograms of those sources (Fig. 20) show that the sources only present in the prior catalog (prior-only sources) mostly lie at the lowest-S/N_{peak} end where the spurious fraction is 50%. The few prior-only sources at high S/N_{peak} are blends with nearby prior sources, such that only one source is cross-matched to the corresponding PyBDSF counterpart. The sources only present in the blind catalog could be spurious (if at low ALMA S/N) or, if at high ALMA S/N, real dusty, high-redshift galaxies whose optical/near-IR/radio emission are too faint to be detected in the prior catalogs. However, as there is currently no optical/near-IR information available for these blind-only sources, we exclude them from the analysis in the rest of this paper.

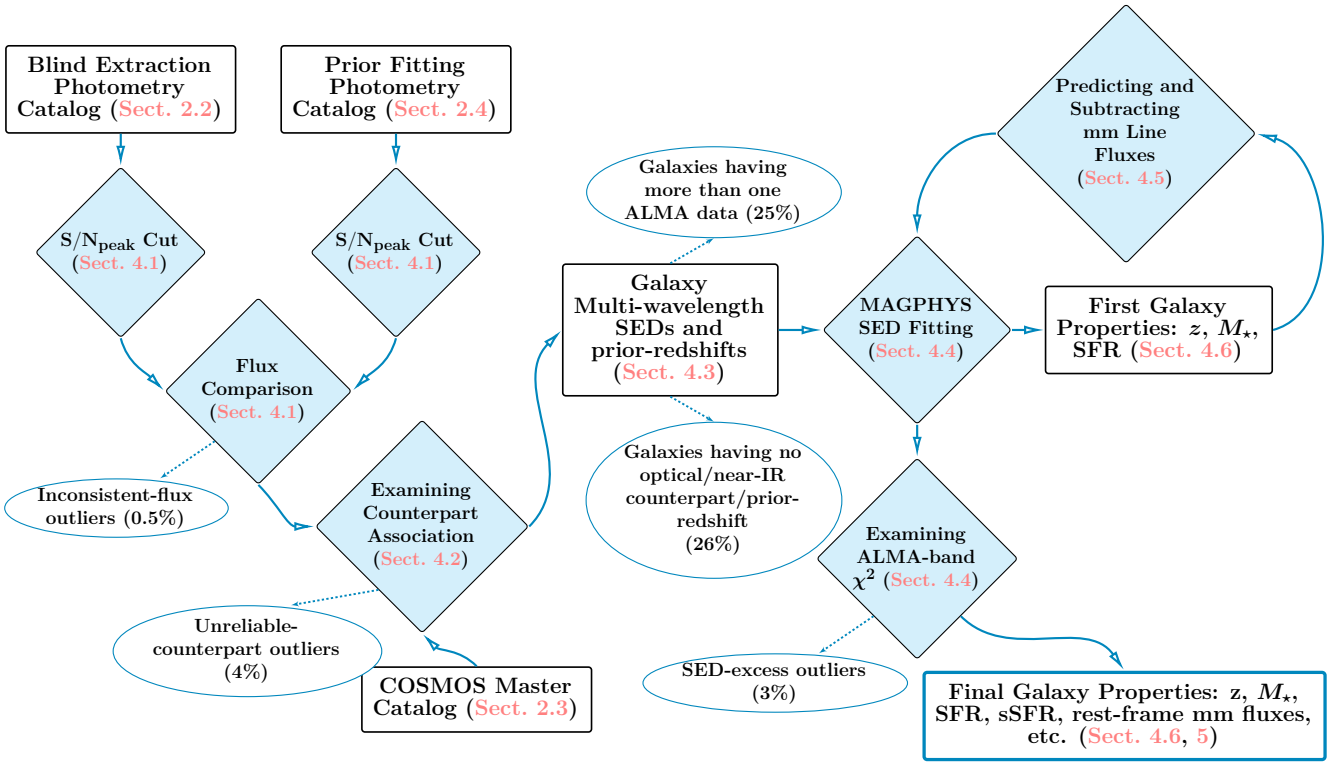

Fig. 18. Work flow for the selection of a reliable galaxy sample and the determination of its properties (see Section 4). We first apply a S/N$_{peak}$ cut to our two photometry catalogs, then apply a counterpart association code (based on machine learning) to construct our galaxy multi-wavelength catalog. Next, we run SED fitting to identify outliers that are either due to spurious ALMA sources or "suspicious" (inconsistent) redshifts in the literature. Finally, after discarding spurious sources and refinement of inconsistent prior redshifts, SED fitting is repeated to obtain physical properties of our galaxies. The corresponding subsection in the text are provided in brackets.

After accounting for 25% galaxies having more than one ALMA observations, either due to different wavelengths or spatial resolutions, we have 823 unique galaxies (with dataset version `20180201`). The ALMA flux densities and their errors are then corrected for the primary beam attenuation. As 26% of these galaxies do not have sufficient optical/near-IR data, i.e., not in Laigle et al. (2016) catalog, it is not possible to obtain reliable stellar masses for them. While some of these sources emit weakly in the deeper IRAC 3.6 μm and 4.5 μm data from the Spitzer Large Area Survey with Hyper-Suprime-Cam (SPLASH) survey (PI: P. Capak; I. Davidzon, priv. comm.) and are also present in the IRAC catalogs from the Spitzer Matching Survey of the UltraVISTA Ultra-deep Stripes (SMUVS; Ashby et al. 2018), their stellar masses and photometric redshifts have large uncertainties due to the lack of shorter wavelength information. We therefore omit these sources from our galaxy catalog (see the "no optical/near-IR prior-redshift galaxies" entry in Fig. 18; they are kept in the ALMA photometry catalogs, e.g., those with IRAC/radio priors). We plan to up-date our galaxy catalog when deeper optical-to-K-band data becomes available, e.g., from the UltraVISTA Data Release 4.

In the next sections, we further exclude some outliers from the ALMA photometry catalogs to construct our final galaxy catalog. We list the numbers and fractions of sources excluded at each step in Table 3.

4.2. *Examining counterpart association*

Our ALMA data set has excellent spatial resolution ($\sim$ 1″) compared to data from single-dish (sub-)mm telescopes ($>$ 10″), and for most sources an unique counterpart at optical/near-IR/radio wavelengths can be easily identified by examining the spatial separation. However, small number of ambiguous cases remain for both prior fitting and blind extraction photometry. Note that we have already corrected for the known astrometry offsets between prior and ALMA positions before our final run of prior fitting (for more details on astrometry, see Appx. A).

During our prior fitting photometry, we allow the source position to vary if the source has high S/N (see Sect. 2.4). This implies that any ALMA source not in our prior master catalog close to a prior position will be

Table 3. Number of Sources in A^{3}COSMOS Catalogs and Excluded at Each Step in Sect. 4 (Version `20180201`; See Also Work Flow in Fig. 18)

Catalog/Step	Number	Fraction
Prior-photometry catalog	1027	···
Blind-photometry catalog	939	···
Combined ALMA detections	1146	···
Galaxies having more than one ALMA data points (Sect. 4.1) [a]	204	25%
Galaxies having no optical/near-IR counterpart/prior-redshift (Sects. 4.1, 4.2 and 4.3) [b, c]	215	26%
Inconsistent flux outliers (Sect. 2.5; `Flag_inconsistent_flux`) [c]	4	0.5%
Unreliable counterpart outliers (Sect. 4.2; `Flag_outlier_CPA`) [c]	36	4%
SED excess outliers (Sect. 4.4; `Flag_outlier_SED`) [c]	21	3%
Final galaxy catalog (Sect. 5) [d]	676	···

[a] In this step, we sorted 1027 ALMA prior detections into 823 unique galaxies, while discarded 119 blind-only sources (see discussion in Sect. 4.8). The fractions in the third column are of the 823 unique galaxies.
[b] This includes the 119 blind-only sources, 43 galaxies which have no redshift from literature as prior information and 53 galaxies which only have a far-IR/mm photo-z from Jin et al. (2018). They are excluded from the further quality assessments due to too poor constraints on galaxy properties.
[c] 10 sources are duplicated among these flags.
[d] Our approach aims at keeping only galaxies with most reliable properties (redshift, stellar mass and dust-obscured SFR), therefore the number of galaxies is significantly reduced compared to the number of ALMA detections. The exclusion of galaxies does not mean they are all not real, but just their properties could not be reliably estimated with current data. Future follow-ups will be needed to explore their properties.

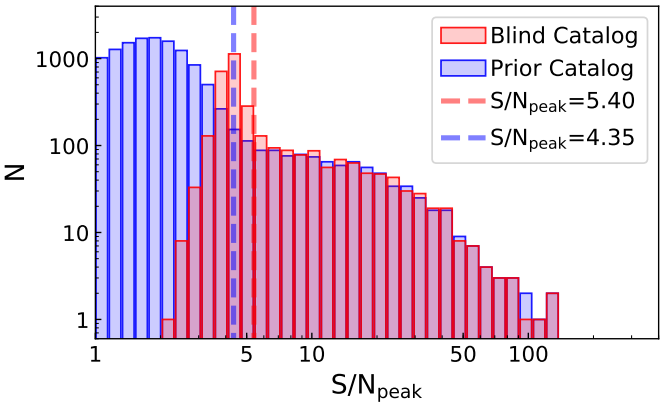

Fig. 19. S/N$_{peak}$ histograms of our blind-extraction and prior-fitting catalogs. The blue and red vertical dashed lines indicate the S/N$_{peak}$ thresholds we applied to select our sample from the prior and blind catalog, respectively.

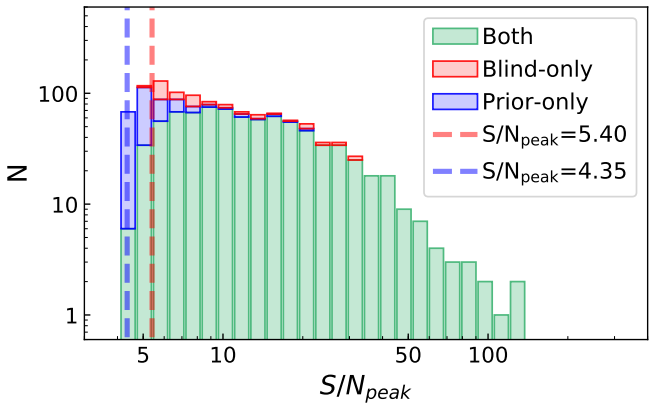

Fig. 20. Vertically-stacked S/N$_{peak}$ histograms of our selected sample from the blind-extraction and prior-fitting catalogs. Sources in both catalogs are indicated by green bars and shown with the prior-catalog S/N$_{peak}$, while those in the blind-extraction (prior-fitting) catalog with S/N$_{peak}$ above the labeled threshold are shown in red (blue) bars. The height of each stacked bar indicates the relative number, and the total height of the histogram represents our selected sample size. (Comparing to Fig. 19, the difference in the third highest bin is due to different S/N$_{peak}$ between prior- and blind-photometry as detailed in Sect. 2.5.)

wrongly attributed to that prior. In these cases, they are more likely to have a certain spatial offset. But this scenario needs to be distinguished from the case where the prior source is an extended galaxy and its dust emission peak is offset from its optical position (e.g., Hodge et al. 2016; Chen et al. 2017).

Besides, spurious sources caused by noise boosting (∼10% spurious sources are expected from our statistical analysis with our selection thresholds in Sect. 4.1) can also exhibit larger offsets, as the signal boosted by noise is randomly spatially distributed. Thus by examining the counterpart association, we can identify most of these outliers (∼4% in this step; or in total ∼ 8.4% including the steps in the next sections) and reduce the number of spurious sources in our final catalog [18].

[18] Note that examining the counterpart association is not helpful in identifying line-of-sight boosting by noise or blending by background source. Therefore the outlier fraction found in this step is only ∼4%, about half of our expected spurious fraction ∼ 8 − 12% (Sect. 4.1). However, as shown in the next section, SED fitting is a powerful tool to exclude ∼ 3% line-of-sight outliers and further reduce the spurious source fraction in our final catalog. In total, after `Flag_inconsistent_flux` (Sect. 2.5), `Flag_outliers_CPA`

In order to correctly identify such ambiguous cases in an automated fashion, we quantify the counterpart association process by several measurable parameters as follows:

- The projected separation between the positions of the ALMA and counterpart source, normalized by the projected ALMA source radius (denoted as **Sep.**);
- The ALMA total flux S/N (denoted as $\mathbf{S/N_{ALMA}}$);
- The S/N of the aperture-integrated flux in optical/near-IR/radio images, measured with an aperture centered at the ALMA source position ($\mathbf{S/N_{S.}}$) and at the reference counterpart positions ($\mathbf{S/N_{Ref.}}$), as well as their respective ratio (denoted as **S./Ref.**). The aperture size is determined via measurements with a series of concentric apertures where the aperture with the maximum S/N is taken;
- An extension parameter **Ext.** which traces the amount of extended optical/near-IR/radio emission within the location between the ALMA and counterpart positions. This is quantified by deriving the optical/near-IR/radio surface brightness level within a series of fixed-size apertures (equal to the fitted ALMA source size) centered along the connecting line between the ALMA position and the reference counterpart position. The linear slope of the relation between surface brightness and increasing (linear) distance from the ALMA position is adopted as **Ext.**: if the source is an extended galaxy and the optical emission is attenuated by dust at the ALMA position, then **Ext.** is around or slightly larger than 1. However if the ALMA source is a dusty galaxy with non-detectable optical emission and is wrongly associated to a counterpart in optical catalog at some distances away, **Ext.** will be very large or even not measurable in the counterpart optical image (as we require $\mathbf{S/N_{S.}} > 3$ in the apertures to measure the **Ext.** parameter).

These parameters have been defined to best describe the counterpart association process, and are best suited to distinguish between those considered true by visual classification from those cases where the visual classification suggests that the ALMA source is unrelated to the counterpart source. These parameters are then measured for each ALMA detection (Sect. 4.1) and its master catalog counterpart (Sect. 2.3) in four counterpart images: HST ACS i-band image from Capak et al. (2007); UltraVISTA K_s-band image from McCracken et al. (2010, 2012); *Spitzer* IRAC 3.6 μm image from the SPLASH survey (PI: P. Capak); and VLA 3 GHz image from Smolčić et al. (2017). Other images have worse spatial resolution and/or sensitivity therefore are less helpful in distinguishing the quality of counterpart associations.

Empirically, we find counterparts with larger **Sep.** and lower $\mathbf{S/N_{ALMA}}$ are less reliable (i.e., less confident to say that the ALMA emission belongs to the counterpart galaxy, based on our visual identification). However, those could be more reliable if we see extended emission between the ALMA and counterpart position (i.e. **Ext.** ~ 1), which could be the aforementioned case where the galaxy's dust emission is offset from its optical emission and has a smooth transition in-between. We show an example of our counterpart association diagnostic in Appx. D.

With these parameters, we proceed with machine learning techniques to establish the linkage between these parameters and the confidence of a counterpart association. To build up a training data set, three team members visually classified all the 1000+ ALMA detections individually. We visually inspected ALMA contours overlaid on ACS i-band, UltraVISTA K_s-band, IRAC 3.6 μm, and 3 GHz images and assigned each source a classification of 1 (robust) or 0 (spurious or incorrect association). We adopt the median classification from the three sets as truth. In order to automate this classification for future data releases, we use the results from visual inspection to train an algorithm that takes as input the parameters described above (**Sep.**, $\mathbf{S/N_{ALMA}}$, $\mathbf{S/N_{Ref.}}$, **S./Ref.** and **Ext.**) calculated for the ACS, Ks, IRAC 3.6μm, and 3GHz cutouts. In addition, we include a flag for **crowdedness** (defined as the density of master catalog sources weighted by a 2D Gaussian with a FWHM of PSF size; see Liu et al. 2018 Eq. 1) and **clean** parameter (defined as the number of master catalog sources within 3″ radius; Elbaz et al. 2011), as they are helpful in identifying extremely blended cases.

For this supervised machine learning task, we use the PYTHON SCIKIT-LEARN package (Pedregosa et al. 2011). For sources with missing parameters, we replace the missing values with the mean of that parameter from the entire sample. Then we randomly select 60% of the sample with visual classifications for training, leaving the final 40% for model validation. After testing a number of different classifiers available in SCIKIT-LEARN, we decide to use the Gaussian Process (GP) classifier, which implements Gaussian Processes for probabilistic classification. Running our trained model on the valida-

(Sect. 4.2) and `Flag_outliers_SED` (Sect. 2.2), we excluded 61 sources as spurious for 727 quality-assessed galaxies. This is basically in agreement with our statistics ($8-12\%$).

tion sample gave an accuracy of $\sim 96.5\%$. For the total sample of 1027 analyzed sources, we find 94% (965) of sources are classified as robust by both the visual and GP classifications. 3% of the sources (32) are classified as not-robust/spurious by both visual and GP classifications (bringing the overall accuracy to 97%). Only 1% of the sources (7) are classified as robust visually but missed by the GP classification. 2% of the sources were classified as not-robust visually but assigned a robust classification by the GP classifier. Reassuringly, the cases where the visual and GP classifications disagree are all borderline cases where the three visual inspectors are also not in full agreement. The model was saved and can be re-used to predict the robustness of counterpart associations for future A^{3}COSMOS runs without the need for visual classification, provided our current training sample is representative of future datasets.

After this automated counterpart association step, 36 sources are flagged as spurious sources (they could potentially be noise-boosted or a co-aligned real dusty galaxy). We flag them by the `Flag_outliers_CPA` column in our final galaxy catalog, and discard them for our further analysis in this paper.

4.3. *Combining multi-wavelength photometry and prior redshifts in the literature*

To combine the multi-wavelength photometric and spectroscopic information for our prior catalog, we adopt the optical/near-IR photometry from the Laigle et al. (2016) catalog, and use the 3″ diameter aperture fluxes to be consistent with Laigle et al. (2016) [19].

Further, we adopt the far-IR/(sub-)mm/radio photometry from Jin et al. (2018). The authors use detailed "super-deblended" procedures following Liu et al. (2018) to overcome the severe source confusion in their far-IR/(sub-)mm data, which is due to the large beam sizes of the *Herschel* and ground-based single-dish far-IR/(sub-)mm telescopes. Their photometry is prior-based, with the prior catalog constructed by combining the Laigle et al. (2016), Muzzin et al. (2013) and Smolčić et al. (2017) catalogs, all of which are also in our master catalog. The "super-deblended" photometry uses the prior information of galaxies' photometric redshifts and SEDs to "freeze" low-redshift sources, and includes the step of blindly extracting sources in the residual images and re-fitting together with initial priors. Therefore sources not in the prior catalog or even coaligned sources at a significantly higher redshift than the prior source have already been reasonably well accounted for (e.g., if prior redshift < 1, its SED will predict a too low far-IR flux and it gets "frozen" during fitting; see details in Liu et al. 2018 and Jin et al. 2018). More complex situations arise if an unknown far-IR source is blending with a prior source whose SED is not constrained well. However, the ALMA data have typically the spatial resolution and sensitivity to distinguish them. In this work, we do find about a hundred ALMA sources not in the prior catalog used by Jin et al. (2018), of which only about ten are blended with a Jin et al. (2018) prior source (within 1″), and their ALMA ~ 1 mm fluxes (< 1 mJy) indicate that they are undetectable by *Herschel* and SCUBA-2. Therefore using the Jin et al. (2018) catalog for far-IR photometry seems appropriate, especially for those with common priors.

For the SED fitting in this work, we first consider a prior spectroscopic or optical/near-IR photometric redshift if available in the literature. Using photometric redshift is motivated by the sufficiently good agreement between photometric and spectroscopic redshifts as demonstrated by Laigle et al. (2016).

In this work, we examine all the spectroscopic and photometric redshifts in the literature listed in Sect. 2. We show the comparison of these redshifts (hereafter prior redshift, or "prior-z") in Fig. 21, where each data point represents a galaxy in our galaxy catalog and has prior-z from both the Laigle et al. (2016) catalog and other catalogs [20]: M. Salvato et al. spectroscopic redshift catalog; Davidzon et al. (2017) photometric catalog for the same UltraVISTA galaxies as Laigle et al. (2016) but with optimized SED fitting for $z > 2.5$ sources; Delvecchio et al. (2017) photometric catalog for radio-detected galaxies; and Salvato et al. (2011) photometric catalog for X-ray-detected AGNs.

The majority of our sample galaxies show good consistency among all available prior redshifts. However we do find several types of outliers: (1) About 14 X-ray-

[19] Laigle et al. (2016) found that the 3″ aperture fluxes lead to better photometric redshift determination and are less affected by uncertainties in the astrometry. See their Sect. 4.1.

[20] To make sure we select common sources in these catalogs, we first do a backward cross-matching from each compared catalog to our full COSMOS master catalog (Sect. 2.3; with 1″ radius). Then we identify common sources by matching the exact master catalog ID. This avoids linking of different sources in the different catalogs which are closer than our cross-matching radius of 1″. While the nominal false-match probability with this matching radius is 11% (applying Eq. 1 of Pope et al. 2006), we note that the it is only indicative of the likelihood of spurious cross matches between catalogs in a statistical sense, based on the number density of sources and distance between counterparts, but does not include physical information about these matches. Since we have a priori information about whether catalog matches are physically realistic, the actual value of the "false-match probability" will be lower than the listed values in this manuscript.

detected AGNs have higher redshifts in the Salvato et al. (2011) than in the Laigle et al. (2016) catalog (see open squares in Fig. 21), but about half (6) of them have spectroscopic redshifts in good agreement with the Salvato et al. (2011) values (see overlap between open squares and yellow circles in Fig. 21). (2) About 25 $z > 3$ galaxies have lower redshifts in Davidzon et al. (2017) than in Laigle et al. (2016), as indicated by the black solid circles in Fig. 21, but about half (14) of them have consistent second redshift peaks in Laigle et al. (2016) (see the black solid circles with white cross in Fig. 21). (3) About 10 low quality spectroscopic redshifts (i.e., with two or less detected spectral features to determine the respective redshift) disagree with Laigle et al. (2016), yet both could have large uncertainties (see yellow circles outside the area enclosed by dashed lines in Fig. 21).

In our next step, we will run SED fitting to obtain galaxies' stellar mass and star formation rate properties, but with redshift fixed to a prior-z [21]. For galaxies with consistent prior-z or a single prior-z from the above catalogs, we directly use it for the SED fitting. But for galaxies with inconsistent prior-z ($\Delta z > 0.15 \times (1+z)$) from the above catalogs, we run SED fitting for each inconsistent prior-z and take the one with minimum-χ^2 at the ALMA bands as our best-fit. The details are presented in the next section.

4.4. *SED fitting*

We use MAGPHYS (da Cunha et al. 2008, 2015) [22] for the SED fitting, as it has rich stellar SED libraries and has been widely tested on local and high-redshift galaxies (e.g., Smith et al. 2012; Berta et al. 2013; Rowlands et al. 2014b,a; Smith & Hayward 2015; Hayward & Smith 2015; Smolčić et al. 2015; Miettinen et al. 2017a,b; Delvecchio et al. 2017; Hunt et al. 2019). It assumes an energy balance between the energy attenuated by dust in the UV/optical and that radiated by dust at IR/mm wavelengths. As it is debated if this energy balance is still robust for very dusty galaxies (e.g., Simpson et al. 2017; Casey et al. 2017), we provide some supporting evidence for the assumption of energy balance in our whole galaxy sample (see below in this section).

Due to the large number of templates being fitted, MAGPHYS per default fits the SED at a fixed prior-z (that can be either photo-z or spec-z from the literature). A wrong prior-z can easily lead to a poor fit with a large residual at the wavelengths of the ALMA bands,

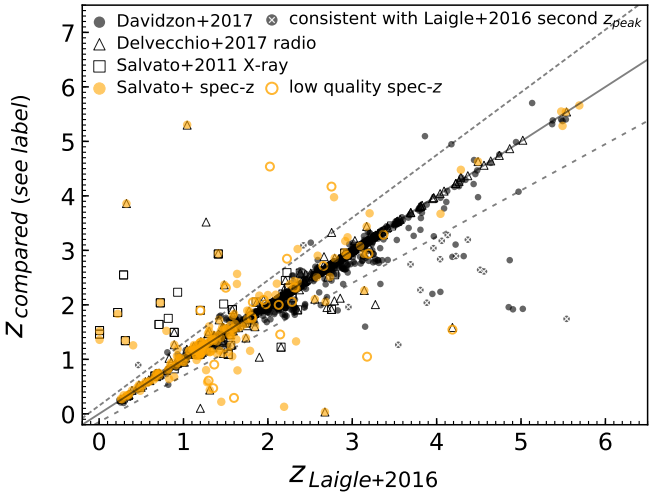

Fig. 21. Comparison between literature photometric and spectroscopic redshifts available for our galaxy sample. Redshifts from various studies in the literature: Davidzon et al. (2017); Delvecchio et al. (2017); Salvato et al. (2011) and the M. Salvato et al. compilation catalog of spectroscopic redshifts are plotted against the photometric redshifts from Laigle et al. (2016). Each data point represents a master catalog source that has a counterpart in the second respective catalog (see Sect. 4.3 footnote 20 for the cross-matching). Solid orange circles indicate sources with robust spectroscopic redshifts ($\geq$ 2 spectral features); low quality spectroscopic redshifts with only one spectral feature are shown as open orange circles. From the set of sources with photometric redshifts in Davidzon et al. (2017), we highlight those that have a consistent second probability peak in redshift in the Laigle et al. (2016) catalog by a white cross inside the black circle (mostly around redshift 4 in Laigle et al. (2016) catalog). The solid line presents the one-to-one relation, and the dashed gray lines indicate $\pm 0.15 \times (1+z)$ catastrophic errors (e.g., Laigle et al. 2016, Sect. 4.3).

which is measured by the reduced chi-square:

$$\chi^2_{\rm ALMA} \equiv \sum_{\rm ALMA} \frac{(|S_{\rm SED} - S_{\rm OBS}|)^2}{\sigma^2_{S_{\rm OBS}}} / N_{\rm ALMA} \qquad (9)$$

where $\sigma_{S_{\rm OBS}}$ is the flux error and $N_{\rm ALMA}$ is the number of ALMA data points. Therefore we consider all possible prior-z's for a given galaxy and fit each of them before choosing the fit with lowest $\chi^2_{\rm ALMA}$ as the final best fit. [23]

The final values of $\chi^2_{\rm ALMA}$ are generally well behaved. In Fig. 22, we compare the difference between $S_{\rm SED}$ and $S_{\rm OBS}$ at all available ALMA bands for each galaxy. The median of $S_{\rm SED} - S_{\rm OBS}$ for all total flux $\rm S/N_{total} > 3$

[21] We have also run another set of SED fitting without a prior-z, which is presented later in the last paragraph of Sect. 4.4.

[22] http://www.iap.fr/magphys/

[23] We treat spec-z's same as photo-z's, except that only when the $\chi^2_{\rm ALMA}$ of a fitting at a spec-z is at least a factor of 1.5 worse than that fitted at a photo-z, we discard the spec-z fitting.

ALMA photometry is consistent with being zero, suggesting that MAGPHYS fitting has no obvious systematic over or underestimation of the flux. There are about 25% of data points with $S/N_{total} < 3$ (but S/N_{peak} meets our sample selection criterion) which are shown as 3-σ upper limits, and 60% of them are consistent with the SED flux (being above the one-to-one line). The histogram of $\log_{10}(S_{OBS}/S_{SED})$ in the lower panel is fitted with an 1D Gaussian with $\mu = -0.01$ and $\sigma = 0.05$. Its upper 5-σ envelope corresponds to $S_{OBS}/S_{SED} = 1.77$, above which we do find 3% outliers. Most of these "SED-excess" outliers have low total flux S/N (i.e., $S/N_{total} < 4-5$ as indicated by the color-coding in Fig. 22).

We speculate that the outliers are most likely spurious sources boosted by noise which by chance align with their optical/near-IR counterparts and are thus not removed by our earlier counterpart association step. Since their S/N_{peak} pass our previous sample selection criterion, they tend to be large in angular size. And this number is actually supported by the statistics: we expect $\lesssim 12\%$ ($\lesssim$140) spurious sources due to our S/N_{peak} selection in Sect. 4.1, which is then reduced by $\sim 4\%$ by our counterpart association examination in Sect. 4.2. Meanwhile, we have ∼130,000 master catalog sources within the current dataset totaling 946 arcmin^2 regardless of primary beam areas (∼23,000 within primary beam areas which sum up 164 arcmin^2); so we expect a false-match probability of 3% with a matching radius of 0.5″ (Eq. 1 of Pope et al. 2006), i.e., only ∼4 spurious sources to coincide with some prior sources by chance alignment.

Although, we note that there is also a chance that there is an unidentified ALMA source at the same line-of-sight as the foreground prior source thereby boosting the ALMA flux to much higher than what SED could fit. These SED-excess outliers are rare but do exist, e.g., the $z \sim 5.7$ background ALMA source "CRLE" found by Pavesi et al. (2018), which is not in any optical/near-IR/radio catalog but is at the same line-of-sight with a foreground $z \sim 0.3$ galaxy in Laigle et al. (2016) catalog.

Similar to the counterpart association flagging, we flag 21 sources as SED-excess outliers. They are indicated by the `Flag_outlier_SED` column in our final galaxy catalog, and will no longer be considered in our further scientific analysis.

Furthermore, in order to verify whether doing a completely blind photometric redshift scan could lead to better fits (smaller χ^2) or not, we adopt the recently developed photo-z version of the MAGPHYS code (MAGPHYS+photo-z; A. Battisti et al. in prep.). It considers redshift as a free parameter between z=0 and 8 and generates identical libraries to the original version of MAGPHYS for each redshift. The output of this step is a probability distribution function (PDF) of the photometric redshift. We perform this photo-z fitting for all our sources and compare the best-fit redshifts (derived as the median of the PDF) to available spectroscopic redshifts, finding no obvious systematic offset (an 1D Gaussian fitting to the distribution of $(z_{photo.} - z_{spec.})/(1 + z_{spec.})$ gives $\mu = -0.015$ and $\sigma = 0.045$). The comparison with all prior-z also shows no obvious systematic offset (an 1D Gaussian fitting to the distribution of $(z_{photo.} - z_{prior.})/(1 + z_{prior.})$ gives $\mu = 0.000$ and $\sigma = 0.076$).

Comparing the physical properties obtained from the two SED fitting for common sources, we find a median difference (scatter) of 0.0 dex (0.05 dex) and 0.0 dex (0.04 dex) for $\log M_*$ and $\log$ SFR, respectively. However, we do note that the uncertainties in $\log M_*$ and $\log$ SFR are systematically larger in photo-z SED fitting when the uncertainties in redshift are included. (The histogram of the difference in uncertainty has a median of 0.0 dex but has a second peak at 0.2 dex and extends to 0.4 dex.) Therefore, for $\log M_*$ and $\log$ SFR in our final galaxy catalog, we take the uncertainties from the photo-z SED fitting which includes the redshift uncertainty, while keep the best-fit values still from the best prior-z fit.

In this photo-z experiment, we also tested the photo-z of the SED-excess outliers, finding that for 7 of them the photo-z are between z=2–4 whereas the prior-z's are below z=1, while the remaining 14 have photo-z and prior-z consistent with z=0–2. Note that the MAGPHYS photo-z fitting places more weight on the stellar SED when the optical/near-IR bands have more data points than the FIR/mm bands. Thus these SED-excess outliers will still show an excess in their observed ALMA fluxes relative to the SED predicted flux. Given their unreliable photo-z's, such sources will benefit from a better FIR/mm coverage as will be available from future submm/mm surveys like JCMT/SCUBA2 S2COSMOS (at 850 μm; PI: I. Smail), STUDIES (at 450 μm; PI: W. Wang), IRAM 30m/NIKA2 Cosmology Legacy Survey (N2CLS; at 1 & 2 mm; PI: G. Lagache), and the LMT/TolTEC Ultra-Deep Galaxy Survey (at 1 & 2 mm).

4.5. *Correcting significant contribution from emission lines*

In sensitive (sub-)mm observations like the majority of the ALMA observations in the COSMOS field, strong (sub-)mm spectral lines like [C II], [N II] and high-J CO emission from high-redshift galaxies can strongly bias the dust continuum measurement if they

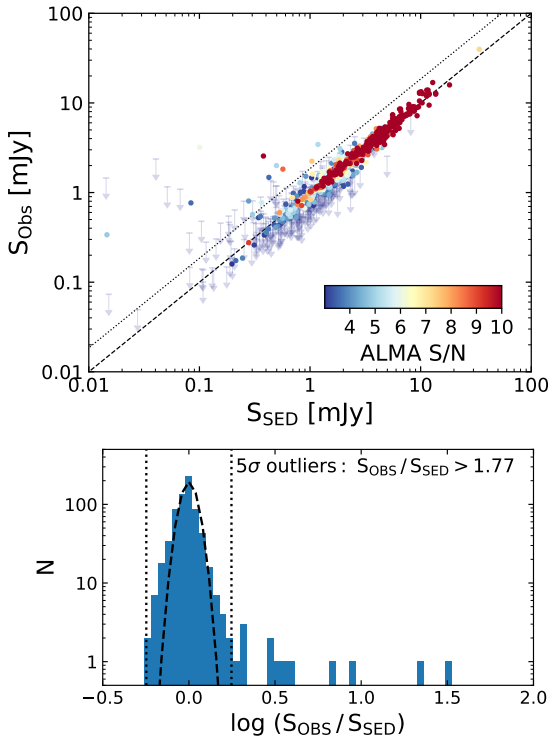

Fig. 22. Upper panel: Comparison of MAGPHYS SED-predicted (S_{SED}) and observed fluxes (S_{OBS}; already corrected for flux bias and error based on our simulation in Sect. 3.1.5) at all available ALMA photometric bands for each galaxy in our sample (Sect. 4.1; removed spurious sources/outliers in Sect. 4.2). Color indicates the S/N of the measured total ALMA flux. Arrows are the 3σ upper limits for sources with total flux S/N < 3. The dashed line shows the one-to-one relation and the dotted line indicates the 5-σ threshold derived from the histogram in the lower panel. **Lower panel:** Histogram of $\log_{10}(S_{OBS}/S_{SED})$ for sources with a total flux of S/N $\geq$ 3 (i.e., excluding upper limits). The dashed curve shows the best fit 1D Gaussian. The vertical dotted lines indicate the 5σ range. We identify sources outside the 5σ range (i.e., $S_{OBS}/S_{SED} > 1.77$) as "SED-excess" outliers (see text for details).

are bright enough and fall in the bandwidth of the spectral set-up [24]. In special cases, these lines will dominate the emission from the whole bandwidth, e.g., mostly ~8 GHz of the current ALMA receiver. This is more significant in the lower frequency 3mm observations, and will become more critical in the future with even deeper observations from ALMA and for the large surveys mentioned in the previous section. It is therefore necessary to consider strong submm/mm line emission together with the dust continuum in photometry pipelines. Although when the observation is not intended for line detection, the chance of a strong emission line being in the bandwidth is very low (e.g., ~ 1.6%, from the blind [C II] line search work by Cooke et al. 2018 who found 10 line emitters out of 695 ALMA continuum sources), but when the number of sources becomes large as in this and future works (with automated pipelines), the line emitters must be systematically corrected for.

As our continuum images are obtained by directly collapsing all channels of all spectral windows ignoring whether the PI intended a line detection or not, a strong (sub-)mm emission line could potentially "contaminate" the measured continuum flux. Therefore we developed a pipeline to automatically identify such cases and to apply a first rough correction for these lines. Direct blanking of channels affected by line emission before construction of the continuum image would require either a good a-priori knowledge of the redshift or dedicated line searches (that are not part of this project) as well as special treatment of each source present in a single pointing. Both aspects not only result in a significant increase in data volume and analysis time required but also in an inhomogeneous dataset. Given the small fraction of potentially affected sources of 7% (see below), our adopted approach is sufficient for our purpose.

Our pipeline uses the redshift and SFR (and IR color, e.g., rest-frame $S_{70\,\mu m}/S_{160\,\mu m}$ from SEDs, when necessary) to predict for each source the low-to-high-J CO (upper level quantum number $2 \leq J_{upper} \leq 10$), [C I] $^3P_2 \rightarrow ^3P_1$ and $^3P_1 \rightarrow ^3P_0$ (at rest-frame 370 μm and 609 μm respectively), [N II] $^3P_2 \rightarrow ^3P_1$ and $^3P_1 \rightarrow ^3P_0$ (at rest-frame 122 μm and 205 μm respectively) and [C II] $^2P_{3/2} \rightarrow ^2P_{1/2}$ (at rest-frame 158 μm). We do not account for other lines in this work because those are predicted to fall outside the frequency range or are generally much weaker. The line prediction follows empirical luminosity–luminosity correlations: [C II]–L_{IR} correlation from De Looze et al. (2011), with a [C II] deficit roughly proportional to $L_{IR}^{-0.335}$ when $L_{IR} > 10^{10}$ L$_\odot$ which fits the data best; [N II]–L_{IR} correlation from Zhao et al. (2013, 2016); CO(1-0)–L_{IR} correlation from Sargent et al. (2014); high-J ($J_{upper} \geq 4$) CO–L_{IR} correlation from Liu et al. (2015); and [C I]–L_{IR} correlation based on the data sets in Liu et al. (2015) and Valentino et al. (2018). For CO $2 \leq J_{upper} \leq 3$ lines, we interpolate the line luminosity using the the CO(1-0)–L_{IR} and CO(4-3)–L_{IR} correlations.

Meanwhile, we obtain the exact frequency setups for each ALMA observation from the ALMA archive, and

[24] For example, ALMA can detect [C II] from a SFR $\lesssim$ 50 M$_\odot$ yr^{-1}, $z \sim 5$ galaxy with $\lesssim$ 30 min on-source time (Capak et al. 2015; or only ~ 2 min if SFR ~ 1000 M$_\odot$ yr^{-1}; Swinbank et al. 2012; Cooke et al. 2018), or high-J CO lines from a SFR ~ 500 M$_\odot$ yr^{-1}, $z \sim 1.5$ galaxy with $\lesssim$ 30 min on-source time (Silverman et al. 2015b).

identify the predicted strong (sub-)mm lines within the frequency setups. We estimate the line contribution to the measured continuum by dividing the predicted line flux by the total bandwidth and compare that to the measured continuum. Our prediction suggest that ∼50 (∼ 7%) sources have (sub-)mm lines contributing more than 20% to the measured continuum. We looked into their data cubes and found that most of them do have line emission as predicted, as all except four have accurate redshift from the M. Salvato spectroscopic redshift compilation. A strong emission line is predicted but not found to be present for only three sources with spectroscopic redshift (A3COSMOS master catalog IDs 1236908, 350733 and 418763) and two with photometric redshift (IDs 339509 and 1236904). Interestingly, two sources (IDs 990180 and 861198) without spectroscopic redshifts from the M. Salvato compilation do show a line detection, and their spectroscopic redshifts are also reported in the literature (Lee et al. 2017; Cassata et al. in prep.). More details of the A3COSMOS line search work will be presented in future papers. Here we have measured those (sub-)mm lines [25] to verify our prediction, and the comparison is presented in Fig. 23 where solid symbols are these A^{3}COSMOS sources. Their measured line luminosity (X-axis) and predicted line luminosity (Y-axis) show good agreement (the dashed lines indicate a factor of 2 range). The pipeline also predicts < 20% line contributions for more sources, but as these lines could not be measured at sufficient S/N in the data cube, they are omitted from the figure.

In Fig. 23, we added 234 line detection with S/N > 3 for CO, [C I], [C II] or [N II] from the literature as follows: Albrecht et al. (2007); Baan et al. (2008); Bauermeister et al. (2013); Bertemes et al. (2018); Capak et al. (2015); Carilli & Walter (2013); Daddi et al. (2015); Lee et al. (2017); Magdis et al. (2017); Magnelli et al. (2012); Pavesi et al. (2018); Saintonge et al. (2017); Silverman et al. (2015c); Spilker et al. (2018); Tacconi et al. (2013); Tan et al. (2014); Yao et al. (2003). SFRs from these works and in addition from Sanders et al. (2003) and Brinchmann et al. (2004) are used for our line prediction. The distribution of $\log_{10}(L'_{\rm line,\,observed}/L'_{\rm line,\,predicted})$ has a mean of 0.07 and scatter of 0.27. Some disagreement can be found at the lowest end where line luminosity $L'_{\rm line,\,observed} \sim 10^8$ K km s^{-1} pc^2. As our current data do not cover this faint regime, improvement is postponed to a future work.

[25] The line search is done in the *uv* plane adapting the methodology of Silverman et al. (2015b); D. Liu et al. in prep.; with CASA and GILDAS.

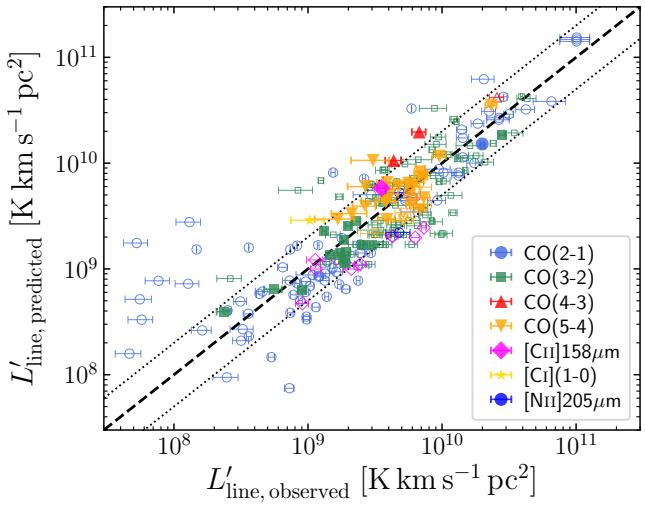

Fig. 23. Comparison of predicted and observed (sub-)mm molecular/atomic line luminosities for a large sample of galaxies with available CO, [C I], [C II] or [N II] luminosity and SFR or IR luminosity in the literature and from this work. This figure verifies our line prediction pipeline which corrects the measured ALMA continuum flux for the emission line "contamination" (see description in Sect. 4.5). Color and symbol indicate different emission lines. Solid symbols are ∼50 sources which have (sub-)mm lines contributing to their measured continuum flux by more than 20% by our prediction. We inspected their data cubes and extracted their (sub-)mm lines and therefore compared to the prediction. Open symbols are 234 galaxies with CO, [C I], [C II] or [N II] detections with S/N > 3 in the literature (see references in Sect. 4.5). The dashed line is a one-to-one line, and the thin dotted lines indicate a factor of 2 scatter.

After the correction for strong (sub-)mm line "contamination", we re-iterate over the SED fitting step. Note that in Fig. 22 the data points represent already the final continuum fluxes corrected for line contamination.

4.6. *Obtaining galaxy properties from SED fitting*

From MAGPHYS SED fitting, we obtain the following galaxy properties: stellar mass (M_*), mass-weighted stellar age, V-band attenuation A_V, star formation history(SFH)-integrated SFR$_{\rm SFH}$, and total IR luminosity $L_{\rm IR}$ (integrated over 8–1000 μm). For each property, MAGPHYS gives a minimum-χ^2 (i.e., best-fit) value, as well as the median and the lower and upper 68$^{\rm th}$ percentiles of the PDF.

Our final SFRs are computed from the IR luminosity with the Kennicutt (1998) calibration and assuming a

Chabrier (2003) IMF:

$$\frac{\mathrm{SFR_{IR}}}{[\mathrm{M_\odot\ yr^{-1}}]} = \frac{L_{\mathrm{IR,\,8-1000\mu m}}}{[\mathrm{L_\odot}]} \times 10^{-10} \qquad (10)$$

By comparing $\mathrm{SFR_{SFH}}$ and $\mathrm{SFR_{IR}}$, we find that the distribution of $\log_{10}(\mathrm{SFR_{IR}}/\mathrm{SFR_{SFH}})$ has more pronounced wings than an 1D Gaussian, with a mean of 0.14 and a standard deviation of 0.15. As mentioned in Kennicutt (1998), the calibration of $\mathrm{SFR_{IR}}$ is based on the starburst synthesis models of Leitherer & Heckman (1995) assuming a constant SFH with a young age of 10–100 Myr (in which time the bolometric luminosity-to-SFR ratio is relatively constant), and assuming that dust re-radiates all the bolometric luminosity. The difference between $\mathrm{SFR_{IR}}$ and $\mathrm{SFR_{SFH}}$ could thus come from either the actual fitted SFHs, the fraction of bolometric luminosity re-radiated by dust, the variation of bolometric luminosity-to-SFR ratio with stellar population ages, or other additional effects. In the following analysis, we will use $\mathrm{SFR_{IR}}$ (and hereafter SFR) because the simple Kennicutt (1998) calibration is widely used in studies focused on the dusty galaxy population at high redshift and given the fact that our sample is biased toward massive, dusty galaxies at high redshift.

Through a detailed simulation and recovery study, Hayward & Smith (2015) tested the accuracy of MAGPHYS in recovering galaxies' physical properties. They found that for isolated disk galaxies, MAGPHYS recovers well the physical properties above mentioned. However, for galaxy mergers, there might be some bias in the determined dust masses (Hayward & Smith (2015) found that MAGPHYS underestimates by 0.1–0.2 dex (up to 0.6 dex) the dust mass during the post-starburst phase of a galaxy merger). Therefore we do not provide dust masses in our final catalog and defer this to later work (D. Liu et al., in prep.).

Hayward & Smith (2015) also found that for AGN host galaxies, when the AGN does not significantly contribute to the UV–mm luminosity (e.g., $< 25\%$), the absence of an mid-IR AGN component in MAGPHYS is not significantly affecting the best-fit results. However, stronger mid-IR AGN can lead to an overestimation of stellar mass and SFR. In our final sample (after removing outliers in Sect. 4.2 and 4.4), 34 galaxies are AGN hosts in the Salvato et al. (2011) *XMM-Newton* catalog and 48 are in the Salvato et al. (2011) *Chandra* catalog. Meanwhile 112 are classified as AGNs via SED fitting with an AGN component using SED3FIT (Berta et al. 2013) by Delvecchio et al. (2017). These catalogs have overlaps, thus the final number of AGNs is 158 ($\sim$23%).

We try to assess the mid-IR AGN problem by running MAGPHYS twice, one time including and the other time excluding the mid-IR 24 μm flux information. Then we adopt the fit with the smaller χ^2 as our final best-fit. The fitting excluding the 24 μm data usually leads to a better χ^2. The overall difference between the derived IR luminosity is very small: the distribution of the difference in $\log_{10} L_{\mathrm{IR}}$ between the two SED fitting results has a median of 0.0 dex and sigma of 0.17 dex. This distribution is slightly broadened to a sigma of 0.25 dex for the AGN subsample, but the median is still close to zero. About 20 sources are 3-sigma outliers, but for most cases the difference is caused by low S/N data at the FIR/mm wavelengths. Only 4 of them are AGNs according to the Delvecchio et al. (2017) classification.

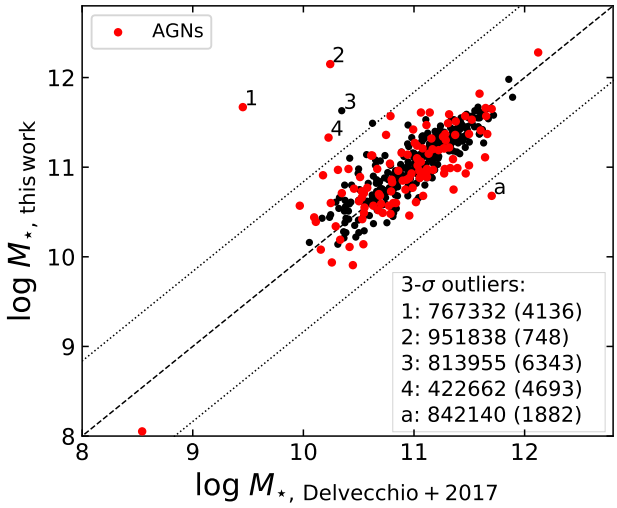

Fig. 24. Comparison of our final M_* from the multi-run MAGPHYS SED fitting and those of Delvecchio et al. (2017) who used SED3FIT to account for the mid-IR AGN component. AGNs classified by Delvecchio et al. (2017) are highlighted in red. Four sources have highly overestimated M_* by our method (labeled `1`, `2`, `3` and `4`), while one sources exhibits a significantly underestimated M_* (labeled `a`). Their corresponding A^{3}COSMOS master catalog ID (and Delvecchio et al. 2017 `ID_VLA3` in brackets) are listed. They are discussed in Sect. 4.6. The dashed line is the one-to-one relation and the dotted lines indicate the $\pm 3\sigma$ range (with σ being the standard deviation).

In Fig. 24 we further compare our final M_* to the Delvecchio et al. (2017) SED3FIT fitted M_* for 396 sources in common (with consistent redshifts and coordinates). AGNs are highlighted in red. This demonstrates a good agreement (within 3σ). We find five outliers (labeled with `1`-`4` if our M_* larger and `a` if M_* is smaller) exceeding the 3σ envelope of the distribution. Their corresponding A^{3}COSMOS master catalog IDs and Delvecchio et al. (2017) IDs are listed in the figure. Through detailed inspection, we find that the difference is mainly caused by including the ALMA data in the SED fitting,

which leads to a higher dust attenuation and thus higher stellar mass.

In addition, the source shown with the highest stellar mass of $\sim 10^{12}$ $M_\odot$ (ID 223951) in Fig. 24 is the strong AGN XID2028 at $z = 1.593$ studied by Brusa et al. (2015), Cresci et al. (2015), Perna et al. (2015) and Brusa et al. (2018). Brusa et al. (2018) estimated a stellar mass of $\log_{10}(M_*/M_\odot) = 11.65^{+0.35}_{-0.35}$ via optical-to-mm SED fitting including an AGN component. For comparison, we obtain $\log_{10}(M_*/M_\odot) = 12.28 \pm 0.07$, almost consistent with their upper boundary. Interestingly, the reduced-χ^2 at the stellar wavelengths of our MAGPHYS SED fitting is as poor as for the outliers 2 and 3 with a `rchi2_star` $\sim$ 6.8 (top $\sim$10% of the worst fits). Delvecchio et al. (2017) accounting for mid-IR AGN contamination obtain $\log_{10}(M_*/M_\odot) = 12.12$ (with an uncertainty of the order of 0.1 dex, see their Sect. 6.1). This indicates that our estimate is still acceptable for such an extreme case (although they should be treated with caution in individual studies).

To summarize our detailed comparison of the robustness of the derived parameters for AGNs, we find:

(1) Our current multi-run, iterative MAGPHYS SED fitting, although without an AGN component, achieves in general good agreements with SED fitting that includes an AGN component. The agreement is valid even for AGN population identified in Delvecchio et al. (2017) and is within the uncertainties even for the most extreme AGNs, e.g., reported in Brusa et al. (2018).

(2) For very few (5 out of 396) sources our stellar masses lie outside the 3σ range when comparing to the Delvecchio et al. (2017) stellar masses. Three of them exhibit strong mid-IR AGN emission contaminating near-IR IRAC and even optical bands. Thus their stellar SEDs are poorly fitted, with `rchi2_star` $\gtrsim 10$. These extreme outliers are further discussed in Appx. E. Their stellar mass estimates in this work should be treated with caution when used in individual studies.

(3) The inclusion of ALMA (and far-IR/(sub-)mm) data points is crucial for codes like MAGPHYS which assumes energy balance. If the energy balance is valid for these dusty, ALMA-detected sources studied here, our stellar masses and IR luminosities are more reliable than optical-only estimates.

4.7. *Final Galaxy Catalog and Properties*

Our final "robust galaxy catalog" contains 676 galaxies with reliable stellar mass and SFR properties, from a parent sample 823 galaxies with at least one ALMA detection in 1534 ALMA archive images (from 142 ALMA projects) available for the COSMOS field (version `v20180201`). In this catalog, 56% of the galaxies have a primary beam correction factor < 1.01 (corresponding to a 2" offset from the phase center at 230 GHz), i.e., they are the primary targets of the PI-led observations. We caution that due to the selection functions of the PI-led ALMA observations in the archive, our sample is not complete in any quantity, e.g., cosmic co-moving volume, stellar mass, and SFR. This bias exists even for sources away from the phase center, because galaxies suffer from clustering effects, and also the input coordinates for single-dish selected sub-mm galaxies are uncertain (possibly resulting in a few arcsec offset from the phase center). Bearing these limitations in mind, we show the redshift, stellar mass and SFR properties of our galaxy catalog in this section and compare with known galaxy correlations and population properties in the literature.

In Fig. 25 we show the distributions of their SFRs and specific SFRs (hereafter sSFR, which is defined as sSFR $\equiv$ SFR$/M_*$, in units of Gyr^{-1}) versus redshift. Our sample spans a large range of SFR from $\sim$1 to $\sim$2000, but the main portion of the sample is SFR-limited at $z > 1$ with SFR $\sim 100 - 1000$.

The low number of $z < 1$ galaxies is mainly due to the selection function, the quick drop of flux density at the Rayleigh-Jeans tail in a galaxy's redshifted SED, as well as the rapid decline of the cosmic star formation rate density at $z < 2$ (e.g., Madau & Dickinson (2014); Liu et al. 2018). Furthermore, given the smaller volume sampled at low redshift, lower source density as well as cosmic variance could play a role as well. Therefore at $z < 1$, our sample is very different from far-IR selected samples (e.g., see a *Herschel* sample in Liu et al. 2018 Fig. 23; see also Béthermin et al. 2015; to name a few). Several $z < 1$ galaxies in our sample are strongly biased toward less massive systems (e.g., $\log_{10} M_*/M_\odot < 10.5$) but also relatively low SFR, and they all have low ALMA S/Ns (total flux S/N $\sim 4 - 5$). Although they passed our rigorous spurious source examinations, they could statistically still be spurious. Should they be real, they are of interest in their own right. However, given these uncertainties we recommend to treat these galaxies with caution especially for individual studies.

In Fig. 26, we show for our A^{3}COSMOS galaxies the distribution of their stellar masses and their SFR offsets to the SFR$_{\rm MS}$ expected for star-forming main sequence galaxies (ΔMS $\equiv \log_{10}$(SFR/SFR$_{\rm MS}$)), where the main sequence SFR$_{\rm MS}$ is defined as a function of redshift and M_* and is empirically measured by a number of works from $z \sim 0$ (e.g., Brinchmann et al. 2004; Chang et al.

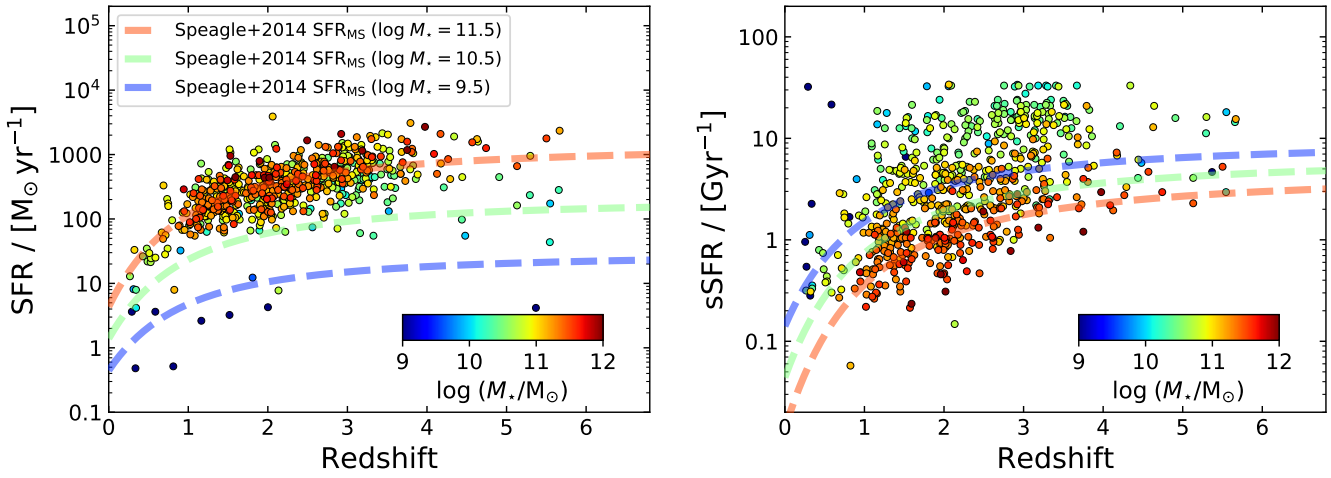

Fig. 25. Distribution of SFR (*left*) and specific-SFR (sSFR) (*right*) versus redshift. Empirical evolution curves of the star-forming main sequence at different stellar masses ($\log_{10}(M_*/\mathrm{M}_\odot) = 9.5$, 10.5 and 11.5, respectively) are shown as blue, green and red dashed lines, which are computed as a function of redshift and M_* following Speagle et al. (2014) (the #49 model in their Table 7). The color bar indicates $\log_{10} M_*$ in both panels.

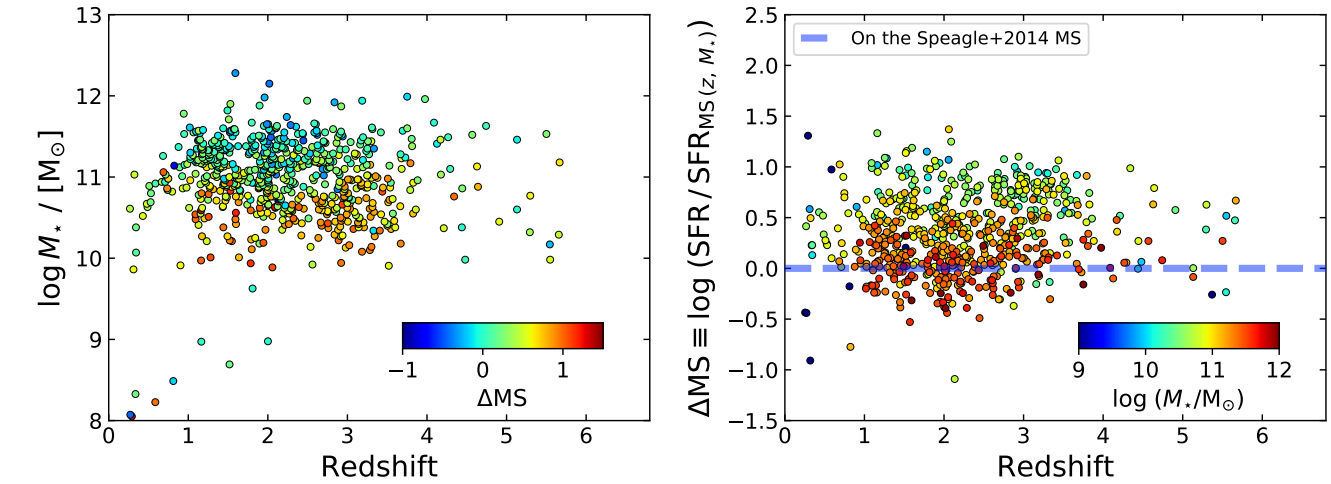

Fig. 26. Stellar mass M_* (*left*) and main sequence offset $\Delta\mathrm{MS} \equiv \log_{10}(\mathrm{SFR}/\mathrm{SFR}_{\mathrm{MS}})$ (*right*) versus redshift, where the main sequence SFR $\mathrm{SFR}_{\mathrm{MS}}$ is computed as a function of redshift and M_* for each source following Speagle et al. (2014) (the #49 model in their Table 7). The color bars indicate ΔMS and $\log_{10} M_*$ in the left and right panels, respectively. The dashed blue line represents $\Delta\mathrm{MS} = 0$ in the right panel.

2015) to ~ 4 (e.g., Speagle et al. 2014; Sargent et al. 2014; Schreiber et al. 2015; Béthermin et al. 2015; Pearson et al. 2018). Here we adopt the Speagle et al. (2014) main sequence (the #49 model in their Table 7).

The majority of our sample lies on the main sequence (i.e., their sSFRs are within a factor of 4, or equivalently ±0.6 dex, of the $\mathrm{sSFR}_{\mathrm{MS}}$; Rodighiero et al. 2011). However, less massive galaxies tend to be above the main sequence. This strong anti-correlation between M_* and ΔMS is primarily an effect of detection limit/sample selection, and is more evident in Fig. 27, where the M_* versus SFR are plotted for A^{3}COSMOS galaxies in nine redshift bins ranging from $z = 0.35 - 5.5$, overlaid with four empirical main sequence parametrizations from Speagle et al. (2014); Sargent et al. (2014); Schreiber et al. (2015) and Béthermin et al. (2015). The differences between these main sequences are relatively small (but see S. Leslie et al. in prep. for a detailed comparison)

The fraction of sources classified as starbursts indicates that our ALMA catalog is biased toward starbursts. It is roughly constant at $\sim 20\%$ in our catalog at each redshift in Fig. 27. But this is a factor of 2–5 higher than that from a *Herschel*-selected sample, e.g., Liu et al. (2018), and much higher than that from a mass-complete sample. For example, Rodighiero et al.

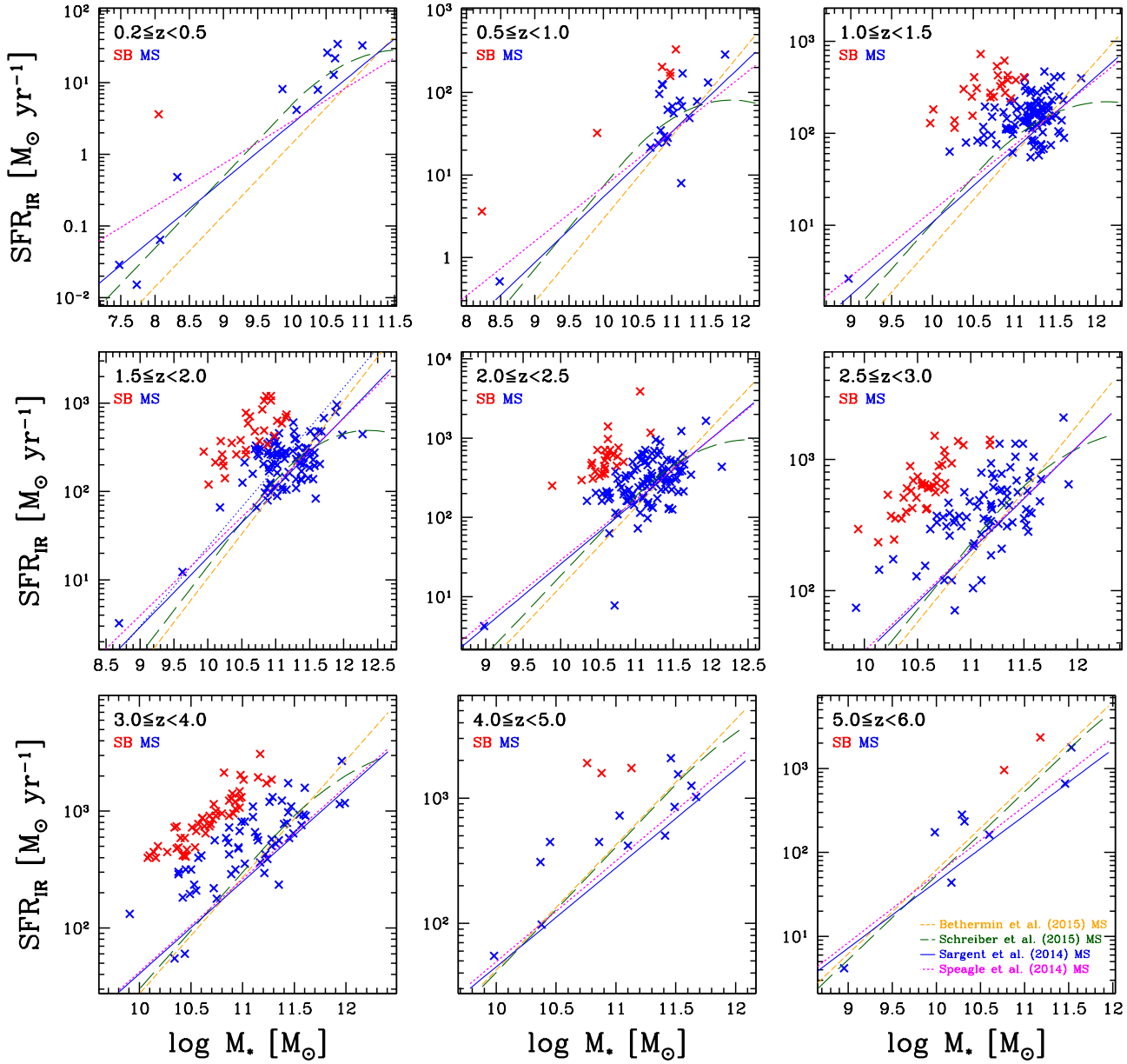

Fig. 27. SFR versus M_*, i.e., the star-forming main sequence diagram, in nine redshift bins from $z \sim 0.35 - 5.5$. Galaxies on the main sequence and starbursts whose SFR is 0.6 dex above the empirical main sequence of Speagle et al. (2014) are shown as blue and red crosses, respectively. Several empirical main sequences from Speagle et al. (2014), Sargent et al. (2014), Schreiber et al. (2015) and Béthermin et al. (2015) are shown for reference (see bottom right panel for information on the color coding).

(2011) find with a starburst fraction of 2%–3% for a sample complete down to $M_* = 10^{10}\,M_\odot$ at $1.5 < z < 2.5$; and Schreiber et al. (2015) report 2–4% for a sample complete down to $M_* = 2 \times 10^{10}\,M_\odot$ and is constant up to $z = 4$. Fig. 27 further shows that it is mainly the less-massive range within which our catalog is dominated by starbursts (e.g., $M_* < 10^{10.5}\,M_\odot$).

Finally, in Fig. 28, we compare the M_* histogram of our sample to the stellar mass functions of star-forming galaxies (Ilbert et al. 2013; Muzzin et al. 2013; Davidzon et al. 2017; Grazian et al. 2015; Song et al. 2016), corresponding to the area of full 2 deg^2 COSMOS field. The completeness of our sample to the full star-forming galaxy population, which can be considered as the fraction of the M_* histogram to the stellar mass functions, strongly depends on redshift and stellar mass. Although the area covered by all the ALMA archival pointings is only about 164 arcmin^2 (or only 4.2% of the full 2 deg^2 area of COSMOS), our sample at $z \sim 1 - 3$ probes a significant fraction ($\sim 10 - 100\%$ depending on the used stellar mass function and redshift bin) of all very mas-

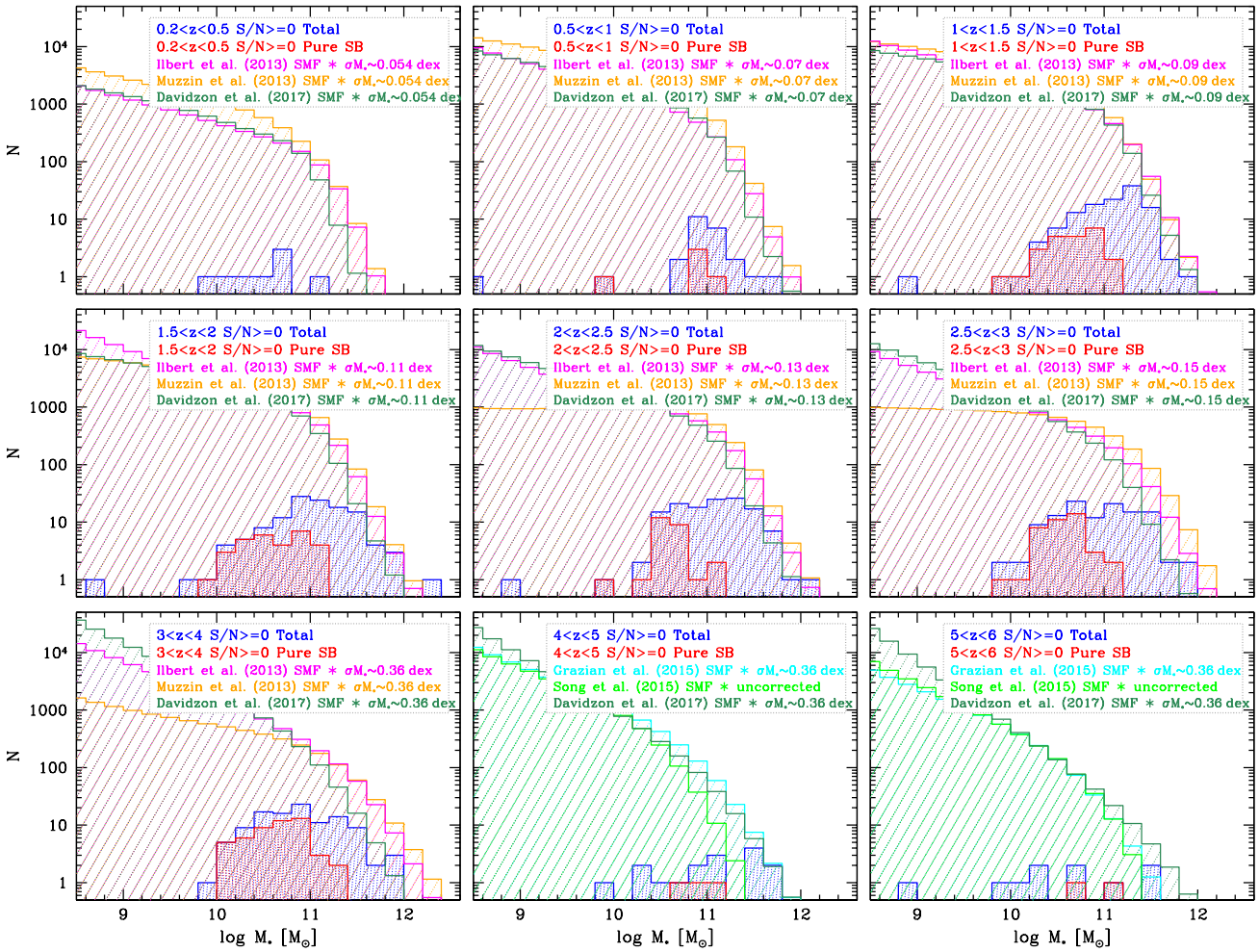

Fig. 28. Stellar mass histograms of our A^{3}COSMOS sources compared to expected stellar mass distributions in nine redshift bins from $z \sim 0.35 - 5.5$. The A^{3}COSMOS sources are selected within a 164 arcmin^2 area covered by 1534 discrete ALMA pointings. A^{3}COSMOS star-forming main sequence galaxies and starbursts as distinguished in Fig. 27 are shown by blue and red histograms, respectively. The colored lines (with line shading) are the stellar mass distributions of all star-forming galaxies within the COSMOS 2 deg^2 area for each redshift bin, computed with empirical stellar mass functions (SMFs) from (Ilbert et al. 2013; magenta), (Muzzin et al. 2013; yellow), (Davidzon et al. 2017; dark green), (Grazian et al. 2015; cyan) and (Song et al. 2016; light green). These SMFs are derived for star-forming galaxies; they have been expressed for a Chabrier IMF when necessary, and convolved with typical stellar mass uncertainties (σ_{M_*}) as indicated by the labels in each panel, following Appx. A of (Ilbert et al. 2013; except for Song et al. 2016, which is the directly observed SMF). The SMFs of Ilbert et al. (2013) and Muzzin et al. (2013) are measured up to $z \sim 4$, while the SMFs of Grazian et al. (2015) and Song et al. (2016) are measured at $z \sim 4 - 6$. Davidzon et al. (2017) SMF probes $z \sim 0.2 - 5.5$ but does not fully cover the highest redshift bin, so we applied a linear extrapolation.

sive ($\log_{10} M_* > 11.5$) star-forming galaxies present in the full 2 deg^2 area.

Due to the large variety of ALMA programs contributing to out dataset, we find no obvious differences between sources at the phase center and in the outer area, even out to a primary beam attenuation of 0.2. The sample selection bias is dominated by the range of sensitivities of the ALMA data rather than the PIs' targeted sources.

4.8. *Properties of the source not included in the final robust galaxy catalog*

As listed in Table 3, a significant number ($\sim 26\%$) of ALMA detections are not included in our final galaxy catalog (see Table 3 caption *b*). Half of them come from the prior-photometry catalog with mostly having only IRAC 3.6 and 4.5 μm and/or VLA 3 GHz priors without optical/near-IR (up to Ks-band) counterparts (hence do not have a photo-z as its prior-z). These "Ks-dropouts" are potential very dusty $z \sim 3 - 4$ galaxies or less dusty sources at even higher redshifts, i.e. similar to the sam-

ple of Wang et al. (2016) and the *HST*-dark sample of Franco et al. (2018). This is in particular true for the sources with significant detections well above our threshold. The remaining half comes from the blind-photometry catalog and has typically low significance implying that they could be spurious, as the differential spurious fraction strongly depends on the actual S/N (see Fig. 8).

As we do not have high spatial resolution optical/near-IR imaging nor accurate photometric redshifts for these sources, it is not possible to do similar counterpart association or SED fitting quality assessments to better identify spurious ones. If we assume that the among of spurious source from quality assessment (about 10% based on the quality-assessed sources) is the same for these 26% sources, then it is 2.6% (all percentages here are of the total ALMA detection number). Adding the two together ($\sim$ 10.1%) gives a total spurious fraction in good agreement with the statistics ($\sim 8-12\%$).

Further discussion of these interesting sources are not the focus of this work. Future deeper optical/near-IR (up to K_s-band) observations, e.g., the new data release of the UltraVISTA survey, will enable an analysis similar to the one done here, so that they could be included in the robust galaxy catalog in the future.

4.9. *The effect of galaxy-galaxy gravitational lensing*

The galaxy-galaxy gravitational lensing has been found to be significant in several ALMA follow-up studies of brightest sub-mm galaxies over large areas, e.g., Negrello et al. (2010), Bussmann et al. (2013, 2015) and Spilker et al. (2016). the strong lensing cases (magnification $\mu > 2$) therein exhibit the following common features: (1) Very bright observed sub-mm flux, e.g., $S_{870\,\mu\mathrm{m}} \gtrsim 15$ mJy for all the $\mu > 2$ galaxies in Bussmann et al. (2013, 2015) and Spilker et al. (2016). (Although we note that lensing is not just limited to the very brightest submillimeter objects but happens at all flux levels, see also below.) (2) Bright optical emission within 1–2" which belongs to a low-redshift (usually $z < 1$) massive galaxy. (3) Usually two or more sub-mm components at each side of the optical emission or roughly distributed as an Einstein ring with $\sim$2" size.

We estimate the number of strongly lensed ($\mu > 2$) cases among our sub-mm galaxies to be very low as follows:

Firstly, given the flux distribution of our photometry catalogs, we only find 0.2% sources with equivalent $S_{870\,\mu\mathrm{m}} \gtrsim 15$ mJy, i.e., four sources in current data set (v20180102). three of them have only very weak or no optical emission in their 1–2″ vicinity, while the fourth one ID 180903 has a low-redshift ($z = 0.347$) optically bright galaxy within 0.5″ and has been already studied in detail by Pavesi et al. (2018). ID 180903 does not exhibit multiple images as expected for strong lenses, fully consistent with its magnification factor of only 1.09 Pavesi et al. 2018.

Secondly, considering the second feature of a close distance to a low-z galaxy, our prior source fitting and SED excess assessment can test for this: if the ALMA flux coming from our prior catalog is originating from a lensed higher redshift galaxy, the SED fitting with a much lower redshift as the prior-z will not be able to fit the ALMA data and therefore be classified as a SED-excess outliers (Sect. 4.4). Among the 21 SED-excess outliers listed in Table 3, we searched for multiple sub-mm images or distorted feature but found no obvious lensed candidates, except for one case ID 650923 where there are three optical components ($z = 0.568$ in Laigle et al. 2016) surrounding the East and South sides of the ALMA emission at a distance of $\sim 1''$ (although the ALMA data has a beam of $1.5'' \times 1.0''$).

Thirdly, there are no multiple sub-mm sources within 1-2" or sources being part of an Einstein ring. This is based on visual identification. In addition, this is confirmed through the comparison between prior- and blind-extraction photometry, which can in principle identify sources with irregular multi-component morphology.

Lastly, our low number of strongly lensed sources is consistent with the analytic galaxy modeling of Béthermin et al. (2017; see also Béthermin et al. 2012b). In their modeling, 1.5 million galaxies are simulated from redshift 0 to 10 within a light cone of 2 deg^2, the same area as the COSMOS field. Their modeled galaxies follow the clustering effect matched to dark matter halos, and strong and weak lensing effect are modeled following Hezaveh & Holder (2011) and Hilbert et al. (2007), respectively. According to our galaxy sample properties, we down-selected 3176 of their galaxies with the criteria $1 < z < 6$, $M_* > 2 \times 10^{10}$ M$_\odot$ and SFR > 200 M$_\odot$ yr^{-1} over the full 2 deg^2. Among this subsample, only 16 have $\mu > 2$. Scaling to our galaxy catalog source number of 823, only three strongly ($\mu > 2$) lensed sources are expected. Note that as discuss in Hezaveh & Holder (2011), there remains significant uncertainty in the estimation of the probability of lensed sources, e.g., the assumed mass model for the lensing halos, the ellipticity of lenses, etc.

Therefore, we conclude that strong lensing is not affecting the properties of most of our galaxies.

5. DATA PRODUCTS

As the results of this work, we produce three public catalogs: two photometric ones (blind-extraction and prior-fitting) and one galaxy catalog (with SED-derived properties). We describe the columns in the first two catalogs in Table 4, and those in the third catalog in Table 5.

The two photometric catalogs have most columns in common, except that the prior photometry catalog has information on the prior source (`ID`, `ID_PriorCat` and `Ref_ID_PriorCat` columns), and some `Flag_*` columns differ.

The `ID` column lists the IDs in our A3COSMOS master catalog which is a combination of 6+ prior catalogs after solving source cross-matching (Sect. 2.3). The `ID` equals to the COSMOS2015 (Laigle et al. 2016) catalog ID when `ID` $\leq$ 1182108. The `ID_PriorCat` column lists the original IDs in those prior catalogs, so that users of our prior photometric catalog can trace back into the prior catalogs. The `Ref_ID_PriorCat` column lists the reference rank number of the prior catalog in Table 2, e.g., the COSMOS2015 (Laigle et al. 2016) catalog has `Ref_ID_PriorCat` = 1, Smolčić et al. (2017) catalog has `Ref_ID_PriorCat` = 5, etc. Note that this reference number indicates in which catalog the source is first included, i.e., has no counterpart in all previous catalogs with smaller `Ref_ID_PriorCat`. Thus our catalog does not contain the information of whether a source with `Ref_ID_PriorCat` = 1 has a counterpart in `Ref_ID_PriorCat` > 1 catalogs (but this information is in our master catalog upon request). Also note that our master catalog will be updated in the future with more deeper prior catalogs, thus we caution that the source ID will be different when future updated master catalog is used.

The `Flag_*` columns carry important information for quality assessment and should be taken into account when using the catalog for specific science applications. For the blind photometry catalog, `Flag_multi` indicates whether the source is fitted with multiple Gaussian components or a single-Gaussian model by PyBDSF. `Flag_inconsistent_flux` indicates whether the source has inconsistent fluxes between prior-fitting and blind-extraction photometry catalog (see Sect. 2.5 and Fig. 7) When `Flag_multi` = 1 and `Flag_inconsistent_flux` = 1, it is likely that the source is a merger system or has a close companion in the ALMA image, as shown in Appx. B. For the prior photometry catalog, `Galfit_reduced_chi_square` indicates the quality of the final GALFIT source fitting. `Flag_size_upper_boundary` indicates whether the fitted source size reaches the upper boundary of 3.0 arcsec we set in the photometry, in which case the source is either blended or dominated by noise and should be used with caution.

The final galaxy property catalog also has two important flags: `Flag_outlier_CPA` which indicates the outliers from our counterpart association examination (Sect. 4.2), and `Flag_outlier_SED` which indicates the outliers with SED-excess from our SED fitting (Sect. 4.4). We recommend to only use galaxies with both `Flag` = 0 for scientific analysis.

The catalogs are available from the COSMOS archive at IPAC/IRSA and in electronic form from the journal. Further the ALMA continuum images are also provided via the COSMOS archive.

6. SUMMARY

The growing information in the ALMA archive is ideal for systematic exploitations of specific scientific questions such as the number and properties of high-redshift galaxies detected in their (sub-)mm continuum emission in selected cosmological deep fields. Given the large number of observations already available in the archive, e.g. for the COSMOS field 1534 pointings covering an area of 164 arcmin^2 have been publically available since Jan. 2nd, 2018, we have developed a highly automatic approach towards mining these data (A^{3}COSMOS — Automated ALMA Archive mining in the COSMOS field). Here we summarize our workflows (Figs. 1 and 18) implemented to obtain quality controlled (sub-)mm source catalogs based on two different identification approaches as well as a catalog of galaxies with (sub-)mm detections and reliable properties.

We present two (sub-)mm continuum source catalogs from public ALMA archival data. For the source identification, the calibrated archival data were homogenously imaged to provide a single continuum image with best sensitivity (i.e. using all available bandwidth and natural weighting). The first catalog is based on a blind extraction using PyBDSF on the continuum images, the second catalog used prior positions from a master catalog that combines sources detected in the optical, IR and radio. Extensive simulations using two mock samples with highly different distributions in (sub-)mm source properties provide robust information on the completeness limits, spurious source fraction, flux boosting factors as well as uncertainties on the measured quantities in both catalogs. In particular, we used these simulations to refine the widely used Condon (1997) prescription for error estimation of radio continuum sources. After further quality control steps the final catalogs (ver-

sion v20180201) contain 939 sources above a peak flux $S/N_{peak} = 5.40$ for the blindly detected sources (with a cumulative spurious fraction of $\sim 8\%$) and 1027 sources above a peak flux $S/N_{peak} = 4.35$ for the prior selected sources (with a cumulative spurious fraction of $\sim 12\%$).

We combine the two (sub-)mm continuum source catalogs to remove inconsistent-flux outliers and use the prior catalog to produce a single sample of high-redshift galaxies with robust (sub-)mm detections by ALMA (25% having more than one ALMA photometric measurement usually at different wavelengths) and mostly homogeneously determined galaxy properties (stellar mass, SFR). The construction included the development of a sophisticated method to automatically qualify counterpart associations with the (sub-)mm continuum sources taking into account astrometric uncertainties (both absolute and relative) as well as complex, differing source structure across wavelength. Further steps were applied to remove spurious (sub-)mm continuum detections and/or sources with highly uncertain redshift information based on SED fitting results. The final galaxy catalog (version 20180201) contains 676 galaxies in the range of $z = 0.25 - 5.67$, $M_* = 3 \times 10^7 - 1.9 \times 10^{12}$ $M_\odot$ and SFR $= 0.02 - 4000$ $M_\odot\,yr^{-1}$. (Despite the vast number of star-forming galaxies presented in this work, we caution that this catalog is not complete in cosmic co-moving volume, stellar mass or SFR.)

The latest version of our catalogs are available from the COSMOS archive at IPAC/IRSA [26] and in electronic form.

DL, PL, and ES acknowledge support and funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 694343). SL acknowledges funding from Deutsche Forschungsgemeinschaft (DFG) Grant SCH 536/9-1. B.G. acknowledges the support of the Australian Research Council as the recipient of a Future Fellowship (FT140101202). Part of this research was carried out within the Collaborative Research Centre 956, sub-project A1, funded by the Deutsche Forschungsgemeinschaft (DFG) – project ID 184018867. We thank Annalisa Pillepich and the Max Planck Computing & Data Facility for very helpful computing cluster resources.

This paper makes use of the following ALMA data: ADS/JAO.ALMA#2011.0.00064.S, ADS/JAO.ALMA#2011.0.00097.S, ADS/JAO.ALMA#2011.0.00539.S, ADS/JAO.ALMA#2011.0.00742.S, ADS/JAO.ALMA#2012.1.00076.S, ADS/JAO.ALMA#2012.1.00323.S, ADS/JAO.ALMA#2012.1.00523.S, ADS/JAO.ALMA#2012.1.00536.S, ADS/JAO.ALMA#2012.1.00919.S, ADS/JAO.ALMA#2012.1.00952.S, ADS/JAO.ALMA#2012.1.00978.S, ADS/JAO.ALMA#2013.1.00034.S, ADS/JAO.ALMA#2013.1.00092.S, ADS/JAO.ALMA#2013.1.00118.S, ADS/JAO.ALMA#2013.1.00151.S, ADS/JAO.ALMA#2013.1.00171.S, ADS/JAO.ALMA#2013.1.00208.S, ADS/JAO.ALMA#2013.1.00276.S, ADS/JAO.ALMA#2013.1.00668.S, ADS/JAO.ALMA#2013.1.00815.S, ADS/JAO.ALMA#2013.1.00884.S, ADS/JAO.ALMA#2013.1.00914.S, ADS/JAO.ALMA#2013.1.01258.S, ADS/JAO.ALMA#2013.1.01292.S, ADS/JAO.ALMA#2015.1.00026.S, ADS/JAO.ALMA#2015.1.00055.S, ADS/JAO.ALMA#2015.1.00122.S, ADS/JAO.ALMA#2015.1.00137.S, ADS/JAO.ALMA#2015.1.00260.S, ADS/JAO.ALMA#2015.1.00299.S, ADS/JAO.ALMA#2015.1.00379.S, ADS/JAO.ALMA#2015.1.00388.S, ADS/JAO.ALMA#2015.1.00540.S, ADS/JAO.ALMA#2015.1.00568.S, ADS/JAO.ALMA#2015.1.00664.S, ADS/JAO.ALMA#2015.1.00704.S, ADS/JAO.ALMA#2015.1.00853.S, ADS/JAO.ALMA#2015.1.00861.S, ADS/JAO.ALMA#2015.1.00862.S, ADS/JAO.ALMA#2015.1.00928.S, ADS/JAO.ALMA#2015.1.01074.S, ADS/JAO.ALMA#2015.1.01105.S, ADS/JAO.ALMA#2015.1.01111.S, ADS/JAO.ALMA#2015.1.01171.S, ADS/JAO.ALMA#2015.1.01212.S, ADS/JAO.ALMA#2015.1.01495.S, ADS/JAO.ALMA#2015.1.01590.S, ADS/JAO.ALMA#2015.A.00026.S, ADS/JAO.ALMA#2016.1.00478.S, ADS/JAO.ALMA#2016.1.00624.S, ADS/JAO.ALMA#2016.1.00735.S. ALMA is a partnership of ESO (representing its member states), NSF (USA) and NINS (Japan), together with NRC (Canada), MOST and ASIAA (Taiwan), and KASI (Republic of Korea), in cooperation with the Republic of Chile. The Joint ALMA Observatory is operated by ESO, AUI/NRAO and NAOJ.

Facility: ALMA

[26] http://IRSA.ipac.caltech.edu/data/COSMOS/

Table 4. Columns in the two photometry catalogs

Column Name	File	Units	Description
`ID`	Prior	—	A3COSMOS master catalog ID (version 20170426), which equals Laigle et al. (2016) COSMOS2015 catalog ID when ID $\leq$ 1182108.
`ID_PriorCat`	Prior	—	The original ID in the `Ref_ID_PriorCat`-th prior catalog.
`Ref_ID_PriorCat`	Prior	—	The reference number of the prior catalog in which the source is first included (see Table 2).
`RA`	Blind+Prior	degree	The fitted R.A. coordinate of the ALMA emission with Gaussian source models, in equatorial coordinate in the epoch of J2000.
`Dec`	Blind+Prior	degree	Same as above but is Dec. coordinate, in equatorial coordinate in the epoch of J2000.
`Total_flux_pbcor`	Blind+Prior	mJy	The fitted total flux with Gaussian source models, corrected for flux bias and primary beam attenuation.
`E_Total_flux_pbcor`	Blind+Prior	mJy	Error in `Total_flux_pbcor`, provided by photometry pipelines based on Condon (1997) simulation statistics and equations.
`E_Total_flux_sim_pbcor`	Blind+Prior	mJy	Error in `Total_flux_pbcor`, but estimated from our own simulation statistics.
`Pbcor`	Blind+Prior	—	Primary beam attenuation factor.
`Primary_beam`	Blind+Prior	arcsec	ALMA 12 meter antenna's primary beam FWHM size at the observing frequency.
`Peak_flux`	Blind+Prior	mJy/beam	Fitted ALMA continuum emission's peak flux, uncorrected for primary beam attenuation.
`RMS_noise`	Blind+Prior	mJy/beam	Pixel rms noise in the continuum image.
`Obs_frequency`	Blind+Prior	GHz	Observing frequency, i.e., the center frequency of all collapsed spectral windows.
`Obs_wavelength`	Blind+Prior	μm	Observing wavelength, = (2.99792458 $\times$ 10^5)/ `Obs_frequency`.
`Maj_beam`	Blind+Prior	arcsec	Synthesized beam's major axis FWHM size.
`Min_beam`	Blind+Prior	arcsec	Synthesized beam's minor axis FWHM size.
`PA_beam`	Blind+Prior	degree	Synthesized beam's position angle, zero means to the North.
`Image_file`	Blind+Prior	—	Image file name.
`Flag_multi`	Blind	—	`Flag` = `S` (or `M`) means the source is fitted with single (or multiple) Gaussian component(s).
`Galfit_reduced_chi_square`	Prior	—	The reduced χ^2 of galfit prior source fitting, measured from the residual image for each source with an aperture of 1.0 arcsec in diameter.
`Flag_size_upper_boundary`	Prior	—	`Flag` = 1 means the fitted source major axis FWHM size reaches the upper boundary of 3.0 arcsec and should be used with caution.
`Flag_inconsistent_flux`	Blind+Prior	—	`Flag` = 1 means the source has $> 5\,\sigma$ inconsistent total fluxes from our prior and blind photometry.
`Flag_outlier_CPA`	Prior	—	`Flag` = 1 means the source is flagged as an outlier in our counterpart association analysis (Sect. 4.2).
`Flag_outlier_SED`	Prior	—	`Flag` = 1 means the source is flagged as an outlier in our SED fitting analysis (Sect. 4.4).

Table 5. Columns in the final galaxy property catalog

Column Name	Units	Description
`ID`	—	A3COSMOS master catalog ID (version 20170426), which equals Laigle et al. (2016) COSMOS2015 catalog ID when ID $\leq$ 1182108.
`RA`	degree	Fitted ALMA continuum emission's R.A. with Gaussian source models, in the equatorial coordinate in the epoch of J2000.
`Dec`	degree	Same as above but is Dec., in the equatorial coordinate in the epoch of J2000.
`z`	—	SED best-fit redshift from the list of prior redshifts in `z_prior`.
`z_prior`	—	Prior redshifts (prior-z) in the literature, multiple values are separated by white spaces.
`Ref_z_prior`	—	References of `z_prior` [a].
`M_star`	$M_\odot$	Stellar mass from our SED fitting at redshift `z`. Assumed Chabrier (2003) initial mass function (IMF).
`L_dust`	$L_\odot$	Infrared 8–1000 μm luminosity from dust from the same SED fitting as above.
`SFR`	$M_\odot$ yr^{-1}	Star formation rate integrated from star formation history from the same SED fitting as above. Same IMF as above.
`sSFR`	Gyr^{-1}	Specific SFR from star formation history, = `SFR` / `M_star` $\times 10^9$.

[a] `A3COSMOS_specz` means the source has the spectroscopic redshift (spec-z) confirmed in our A3COSMOS data cube analysis with at least one S/N $>$ 6 spectral line (Sect. 4.5; Liu et al. in prep). `Salvato2017_specz` means the source has spec-z in the COSMOS spec-z catalog compiled by M. Salvato et al. (available in the COSMOS collaboration; version 07SEP2017 with 103,964 rows). `Salvato2011_Chandra_photoz` means the source has photometric redshift (photo-z) (optimized for AGNs) in Salvato et al. (2011) *Chandra* source catalog and the photo-z is inconsistent with any previous redshift (by $> 0.15 \times (1 + z)$ difference, same condition afterwards). `Salvato2011_XMM_photoz` means the source has photo-z (optimized for AGNs) in Salvato et al. (2011) *XMM-Newton* source catalog and the photo-z is inconsistent with any previous redshift. `Laigle2016_photoz` means the source has photo-z in the COSMOS2015 catalog provided by Laigle et al. (2016) and the photo-z is inconsistent with any previous redshift. `Davidzon2017_photoz` means the source has photo-z (optimized for $z > 2.5$ sources) in Davidzon et al. (2017) catalog and the photo-z is inconsistent with any previous redshift. `Delvecchio2017_photoz` means the source has photo-z (considered mid-IR AGN component) in Delvecchio et al. (2017) catalog and the photo-z is inconsistent with any previous redshift. `Jin2018_photoz` means the source has photo-z (with far-IR/mm photometry) in Jin et al. (2018) catalog and the photo-z is inconsistent with any previous redshift.

APPENDIX

A. ASTROMETRY ACCURACY BETWEEN PRIOR CATALOGS

Variations in the absolute astrometric calibration between different catalogs can cause small, but noticeable offsets between source positions at different wavelengths. As we use prior positions from sources selected from catalogs covering the optical to radio regime, it is important to verify that potential offsets are small. Here we report astrometric offsets between the prior catalogs used from the literature (Sect. 2.3) and our ALMA prior fitting photometry catalog. These astrometric offsets between the prior positions and the fitted ALMA positions are small ($< 0.1''$).

In Fig. A.1, we plot the offsets in R.A. and Dec. for sources common in two catalogs using the UltraVISTA/-COSMOS2015 catalog (Laigle et al. 2016), the VLA-COSMOS 3GHz catalog Smolčić et al. (2017), the HST/ACS i-band (Capak et al. 2007) and the fitted positions of the (sub-)mm sources in our prior-based ALMA catalog (see Sect.2.4). First we confirm that the ALMA astrometry is indeed excellent (as expected for a (sub-)mm interferometer at the angular resolutions and frequencies of our observations) by comparison to the positions of 699 VLA-COSMOS 3GHz sources (*top right panel*). Comparison between the UltraVISTA and our ALMA positions for 827 sources (*top left panel*) yields a a relatively large offset of $+0.10''$ in R.A., but the offset in Dec. is very small ($+0.01''$). We confirm this astrometric offset of the UltraVISTA catalog by examining positions for 9373 sources in common with the VLA-COSMOS 3GHz catalog (*bottom left panel*). Given the order of magnitude larger number of sources the offset of $+0.088''$ in R.A. is statistically meaningful and consistent with the offset seen between UltraVISTA and ALMA source positions. Finally, comparison between positions of 7369 sources in common in the HST/ACS i-band and VLA catalogs (*bottom right panel*) yields a lower significantly offset in R.A. but a more substantial offset in Dec.

Since we allow the source position to vary by a relatively large amount ($\sim 0.7''$, see Sect. 2.4) during the prior-based detection of (sub-)mm continuum source, it is not necessary to repeat the initial detection step. However, we have applied a correction to take the small offsets into account during our counterpart association process (Sect. 4.2). We note that Smolčić et al. (2017) report astrometric offsets of similar size between VLA-COSMOS 3GHz and UltraVISTA source positions using a more complex analysis identifying variations in the astrometry across the full COSMOS field (see their Appx. A.1 and Fig. A.1). As the numbers of sources analyzed per R.A. and Dec. bin are only a few hundred, we prefer to apply only a single value when correcting for the astrometric offset of the UltraVISTA sources.

We note that a new COSMOS photometry catalog is under construction using the UltraVISTA DR4 data which are astrometrically corrected using GAIA data, providing a much better astrometry of a few milli-arcsecs (see https://calet.org/). Our next A^{3}COSMOS updates will use it when available.

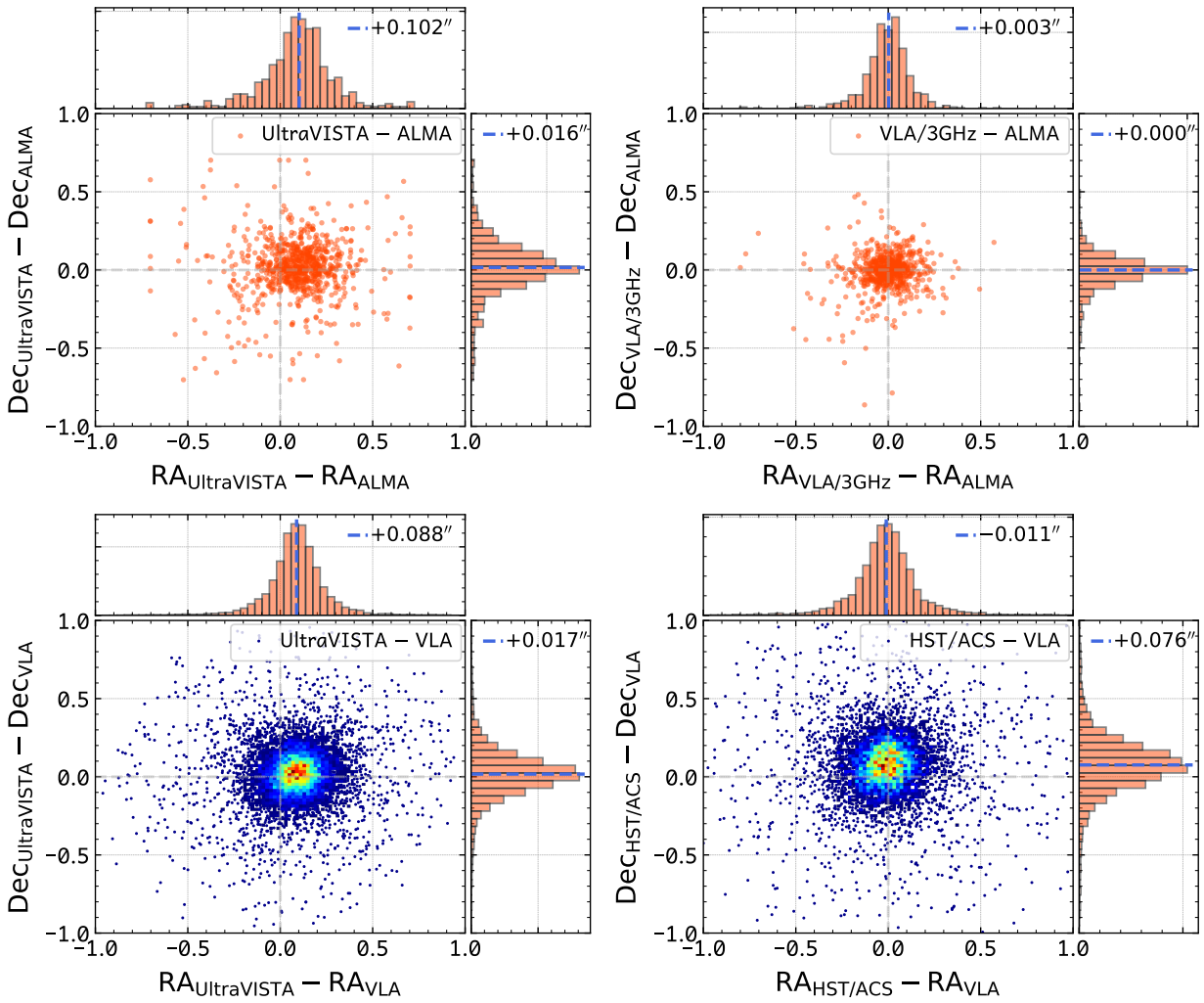

Fig. A.1. Astrometric offsets between source positions in the prior catalogs used from the literature (Sect. 2.3; Laigle et al. 2016; Smolčić et al. 2017; Capak et al. 2007) and positions from our ALMA prior-based detections. Each panel shows the R.A. and Dec. offset distribution for sources common in two catalogs as labeled. Histograms of the R.A. and Dec. offsets are shown at each top and right axis, respectively. Median values of the R.A. and Dec. offsets are indicated by the dashed blue line and the text therein. See Appx. A for the details.

B. SOURCES WITH INCONSISTENT (SUB-)MM CONTINUUM PHOTOMETRY IN OUR TWO CATALOGS

Here we present the ALMA images of the (sub-)mm continuum sources that have inconsistent total fluxes in the prior-based and blindly extracted catalogs and are labeled in Fig. 7. Three outliers with labels `1` to `3` have $> 5\sigma$ higher GALFIT fluxes than PyBDSF fluxes, and one outlier with label `a` has the opposite situation. Their ALMA images, prior fitting and blind extraction model images and residual images are shown in Fig. B.1 (each outlier has six sub-panels; see caption for details).

In general the GALFIT source models provide better fits to the original ALMA images, with less residual emission in the residual images, except for outlier `3` which seems to be composed of two ALMA sources while our COSMOS master catalog contains only one prior source.

For the fourth outlier with label `a` in Fig. 7 and shown in the bottom-right of Fig. B.1, the GALFIT model is more complex than the simple Gaussian shaped PyBDSF model because multiple prior sources are fitted. Therefore, as long as we have a good knowledge of prior sources in the ALMA field of view, i.e., from our compiled COSMOS master catalog, the GALFIT fitting typically provides very good photometry results (with a small enough reduced-χ^2 in the residual image).

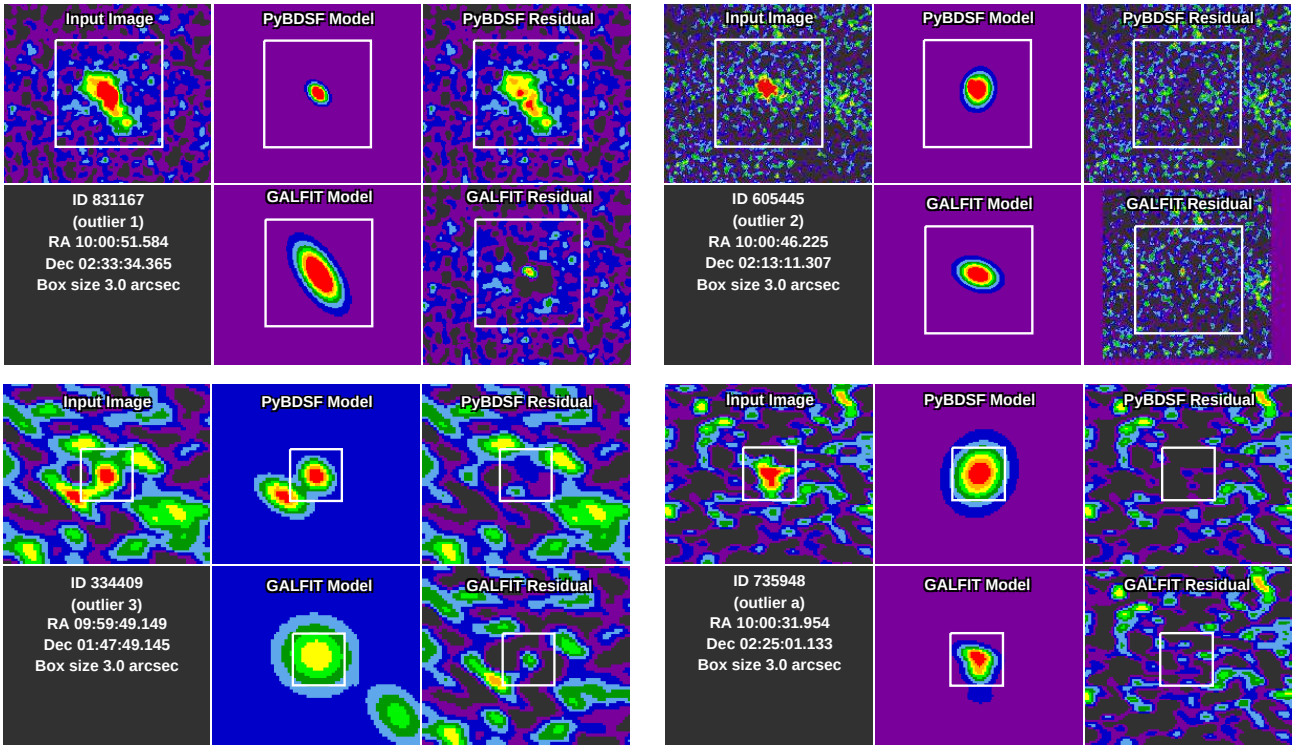

Fig. B.1. PyBDSF and GALFIT fitting images for the four outliers labeled as `1`, `2`, `3` and `a` in Fig. 7 where the measured (sub-)mm continuum fluxes from our PyBDSF and GALFIT photometry differ significantly ($> 5\sigma$. Their IDs in our master catalog (being the same as in the Laigle et al. (2016) catalog) are 831167, 605445, 334409 and 735948, respectively. For each source six sub-panels are shown, namely input image, PyBDSF model image, PyBDSF residual image (*top row, left to right*), source information, GALFIT model image and GALFIT residual image (*bottom row, left to right*). The color scale is the same for all sub-panels for each source, but varies from source to source. A white box with a size of $3''$ is shown in each sub-panel for reference.

C. DETAILED DESCRIPTION OF OUR MONTE CARLO SIMULATIONS

We briefly introduced our two sets of Monte Carlo simulations in Sect. 3 — the full-parameter-space ("FULL") simulation and the physically-motivated ("PHYS") simulation. Below we provide in-depth details of how we model the artificial sources (Appx. C.1.1 and C.2.1), inject them into ALMA residual images (after blind extraction photometry)

(Appx. C.1.1 and C.2.2), recover the sources with our two types of photometry pipelines (Appx. C.1.2 and C.2.3), and analyze the statistics (Appx. C.3). We also discuss the limitations of each simulation in Appx. C.1.3 and C.2.4. Note that both simulations have limitations which could bias our final flux and error estimations. Only by doing both simulations and comparing them with each other as done here, these limitations can be understood and the least biased way to implement corrections to obtain final photometry results can be identified.

C.1. *Full parameter space ("FULL") MC Simulation*

C.1.1. *Source simulation and injection*

In the "FULL" simulation, we simulate one source a time for each of the ∼150 representative ALMA continuum images (Sect. 3), with source peak flux density ranging from 3.0 to 100 times the rms noise (the ratio is denoted as S/N_{peak}; see Eq. 1), and size (Gaussian major axis FWHM, convolved with the beam) ranging from 0.1 to 6.0 times the synthesized beam size (clean beam, Gaussian major axis FWHM) (the ratio is denoted as Θ_{beam}; see Eq. 2).

There are 13 grid points in the first parameter (S/N_{peak}) and also 13 in the second parameter (Θ_{beam}). For each grid point, we generate 25 mock sources by randomizing the injecting position.

The simulated source is assumed to be of Gaussian shape (the minor axis FWHM is generated with an axis ratio randomly picked between 0.2 to 1.0). Then, the source is convolved with the clean beam and injected into the residual image derived from PyBDSF where sources were already blindly extracted and removed. We randomly cut a box area around the source with a size of 8 times the intrinsic source size to ensure enough empty sky area for source extraction.

In total, we have ∼4225 "FULL" simulations per ALMA image, and repeating this for ∼150 representative ALMA images (one for each independent ALMA scheduling block) we have ∼3750 sources per grid point in the two-dimensional parameter space.

C.1.2. *Source recovery*

We run our PyBDSF and GALFIT photometry tools to recover those simulated sources one by one. For PyBDSF, we keep the exact same conditions as for the real catalog, i.e., setting the background to zero and the rms noise to the values we measured from the previous photometry run (from fitting the pixel histograms; Sect. 2.2), and using the same thresholds as for the original ALMA images (Sect. 2.2). For GALFIT, the only difference is the input prior catalog. We assume no source blending issue and only fit the simulated source.

In Fig. C.1 we show the comparisons of the simulated and recovered fluxes for PyBDSF (*top panels*) and GALFIT (*bottom panels*). The left panels show the simulated versus recovered fluxes, colored by S/N_{peak}. Here we only shows sources which have $S/N_{peak} > 3$, because lower S/N detections are mostly spurious, according to the spurious fraction analysis in Sect. 2.8, and eventually we select our ALMA detections with a much higher $S/N_{peak} \sim 5$ (Sect. 4.1).

The middle panels of Fig. C.1 show the difference between the simulated and recovered fluxes ($S_{sim.} - S_{rec.}$) normalized by $S_{sim.}$ as a function of S/N_{peak} (source peak flux to rms noise ratio; Eq. 1), colored by Θ_{beam} (source area to beam area ratio; Eq. 2). In general, at a low S/N_{peak}, $S_{sim.}$ is always smaller than $S_{rec.}$, indicating that fluxes are boosted by noise. Such a flux-boosting is much smaller for a higher S/N_{peak}. Therefore, based on these, we quantify the flux bias by the two parameters S/N_{peak} and Θ_{beam} in the main text (Sect. 3.1). Meanwhile, the scatter of ($S_{sim.} - S_{rec.}$) reflects the uncertainty of the photometry, i.e., flux errors, which can also be quantified by the two parameters (Sect. 3.1.3). Note that the flux errors that came along with our two photometry pipelines are based on the equations in Condon (1997), where the author used about 3000 MC simulations to calibrate these equations. Our MC simulations offer the possibility for alternative assessments that show a broad consistency but also evidence for a second order trend with S/N_{peak} (Sect. 3.1.3 and 3.1.4).

The right panels of Fig. C.1 show the histogram of ($S_{sim.} - S_{rec.}$) normalized by the flux errors $\sigma_{S_{rec.}}$. Sources are grouped in to subsamples according to their S/N_{peak}. An 1D Gaussian fit ($e^{-(x-\mu)^2/2\sigma^2}$) to the histogram indicates whether the flux errors can statistically represent the uncertainty of the photometry. If the fitted Gaussian is too wide (i.e., $\sigma > 1$), then the flux errors are underestimated, and vice versa. We overlay the $\sigma = 1$, $\mu = 0$ 1D Gaussian curve for comparison. Note that both flux bias and error affect these histograms. We demonstrate in Appx. C.3 that after correcting flux biases and re-estimating flux errors, these histograms becomes much more close to $\sigma = 1$, $\mu = 0$ 1D Gaussian shapes, i.e., we can say that they follow a well-behaved Gaussian statistics.

C.1.3. *Limitations*

We discuss several limitations related to the use of "FULL" simulation to indicate flux bias, error and completeness in this section. First is the assumed source property distribution. Using uniform S/N_{peak} and Θ_{beam} distributions

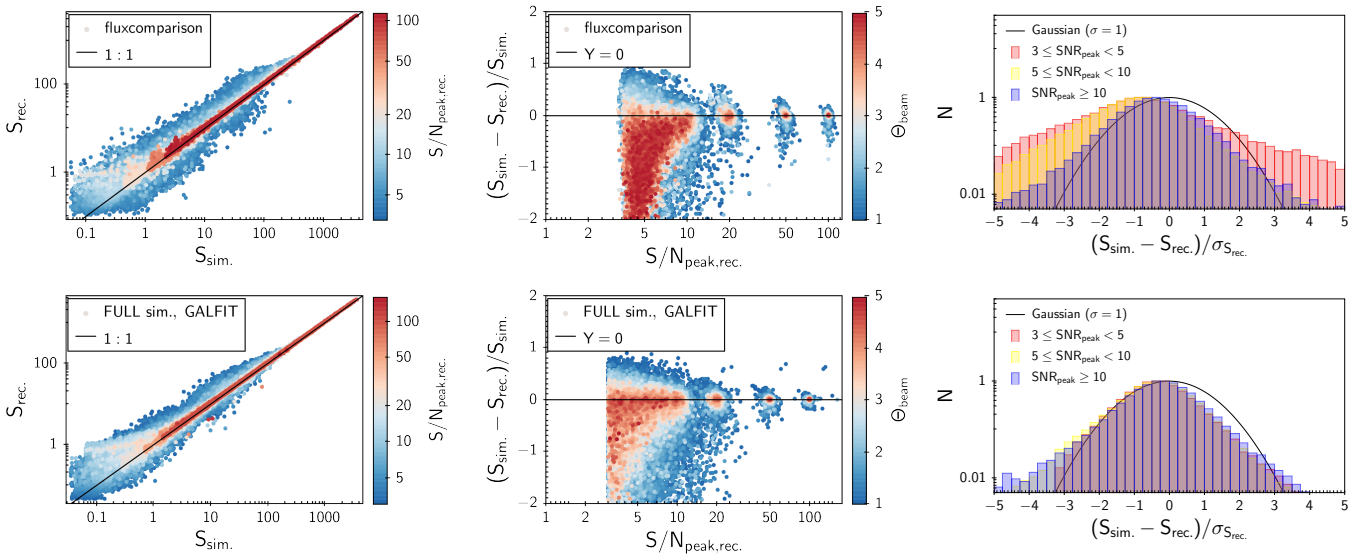

Fig. C.1. Analysis of the PyBDSF (*top*) and GALFIT (*bottom*) recovery of properties of the "FULL" simulation sources. **Left panels** show the comparison of simulated versus recovered flux colored by S/N_{peak}. **Middle panels** present the relative flux difference ($(S_{sim.} - S_{rec.})/S_{sim.}$) versus S/N_{peak}. **Right panels** show the histogram of the normalized flux difference ($(S_{sim.} - S_{rec.})/\sigma_{S_{rec.}}$), where $\sigma_{S_{rec.}}$ is the PyBDSF output flux error following the Condon (1997) equations. Compared to also our corrected histogram in Fig. C.3.

is indeed a strong assumption, although it is perhaps the most commonly adopted way in IR/mm/radio photometry studies. We have to consider the following situation, which we refer to as the *"resolution bias"*. A large, low S/N source can usually break up into several smaller clumps due to noise fluctuations. Our photometry code will then usually only detect a smaller, low S/N clump, therefore the source's total flux is only partially recovered. This acts opposite to the effect of flux-boosting, where our photometry code detects a low S/N source which is actually the peak of a noise fluctuation instead of a real galaxy. In reality, what we know about the detected sources are only the recovered fluxes and sizes, therefore we can not distinguish the two effects. As we parametrize the flux biases and errors by the recovered fluxes and sizes (as will be described in detail in Sect. 3.1), simulating more large size sources will lead to less flux-boosting (hence smaller flux biases), and larger flux errors that can be significant particularly at low S/N.

A second limitation is that sources are simulated and then recovered individually in our "FULL" simulation procedure. Thus there is no source blending or clustering effect. In reality, sources can be blended even at the arcsec resolution of the ALMA data, although this situation occurs at low probability, e.g., it depends on the galaxy merger fraction. This limitation could affect our estimation of the completeness of the PyBDSF photometry, because PyBDSF is a blind extraction tool and sometimes will extract two blended sources as a single source. The GALFIT photometry should not be affected, if the prior source catalog has a high resolution (sub-arcsec) and is complete (not missing sources in ALMA bands).

Another limitation is the input map we used for injecting source models. We use the PyBDSF residual maps for all of our analysis presented here, because the residual maps ideally should contain the exact noise as in the observations. However, imperfect source subtraction (by PyBDSF) could potentially increase the noise at certain positions. However, this is likely a very minor problem as we randomize the injection positions. Injecting source models in the *uv*-plane (pure interferometry noise) instead of the image plane (PyBDSF residual image) could help to assess the additional uncertainty introduced by the imaging/cleaning process. Since we have verified that, for the real data, the rms noise measured from *uv*-plane and image plane are fully consistent, plus the good agreement between the fluxes measured from the *uv* and image planes, both support that this additional uncertainty should be negligible.

C.2. *Physically-motivated ("PHYS") MC Simulation*

C.2.1. *Source simulation*

In order to disentangle the major limitations from the "FULL" simulation, we have done another physically-motivated simulation ("PHYS" simulation) where we try to reproduce the real physical properties of galaxies across cosmic time.

In detail, we follow the two-star-formation-mode (2SFM) recipe (Sargent et al. 2012, 2014; Béthermin et al. 2012a), which assumes all star-forming galaxies are in two populations, with the population of starbursts being enhanced in their sSFRs by a range of factors under a normal distribution with a mean of 5 (here we adopt 5 because we find this better fits the millimeter number counts; Sargent et al. 2014 suggest a value of 4; see their Fig. 10). We generate these star-forming galaxies within the cosmic volume of the 2 square degree COSMOS field with the following procedures:

1) Defining 25 redshift bins from z=9.75 to 0;
2) Computing the number of star-forming galaxies using the stellar mass function (e.g., Davidzon et al. 2017) at each redshift and starting from $M_* = 10^{8.0}$ $M_\odot$;
3) Assuming a small fraction of these star-forming galaxies are in starburst (SB) mode while the rest are in MS mode. The fraction is set to be consistent with the merger fraction extrapolated from Conselice (2014);
4) Computing SFRs for MS and SB galaxies following the MS correlation (including the scatter) and SB boost function in Sargent et al. (2014);
5) Then we generate an IR-to-radio SED for each model galaxy according to the redshift, stellar mass and SFR, following the SED modeling in Liu et al. (2018), which is based on Magdis et al. (2012), assuming Draine & Li (2007) dust models and simplifying the dust SEDs by associating them only to redshift and the interstellar radiation field ($\langle U \rangle$; see Magdis et al. 2012, Béthermin et al. 2015 and Liu et al. 2018 for more details);
6) Estimating MS galaxies' sizes depending on their redshifts and stellar masses following (extrapolating from) van der Wel et al. (2014) as well as considering that dust sizes are a factor of about 2 smaller (Fujimoto et al. 2017);
7) Random minor/major axis ratio from 0.2 to 1;

In total about 0.7 million model galaxies are generated in 2 square degrees. As a validation, their number counts are also estimated at each IR/mm/radio wavelength, these simulated counts are found to agree well with real (sub-)mm measurements at 500 μm, 850 μm, 1.1 μm and 1.3 μm (within the error bars), e.g., Béthermin et al. (2012b), Karim et al. (2013), Carniani et al. (2015), Hatsukade et al. (2016) and Geach et al. (2017).

These model galaxies are then randomly injected into the 2 sq. deg. COSMOS field and recovered, which is described in the following.

C.2.2. *Source placement*

We assign a random position within the 2 square degree COSMOS field for each mock galaxy. Then, we create artificial ALMA maps by inserting our mock galaxies into the ALMA residual images of the 150 representative programs we have selected in Section C.1. For each of the residual maps, we create 273 of such artificial maps. Within each iteration, we select a mock galaxy at specific redshift and stellar mass out of the full mock galaxy catalog, and place it in the center of the residual map. We apply a small random offset (1–6″) to avoid imperfect source extraction at the center of the residual image. A subset of other remaining galaxies from the full mock catalog may fall within the same residual map according to their position within the full simulated 2 square degree map and are inserted as well. In this way, we account for possible clustering of (sub-)mm sources in real observations. We loop for each simulated galaxy over a redshift grid ranging from 1.0 to 6.0 in steps of 0.25, with log stellar mass from $\log(M_*(M_\odot))$ =9.0 to 12.0 in steps of 0.25, making a total of 273 iterations. Then, we extract all the sources simultaneously in the next section (unlike for the "FULL" simulation where we extract each simulated source individually, see Appx. C.1).

C.2.3. *Source recovery*

The simulated images are then treated by our PyBDSF and GALFIT pipelines in the same manner as the original ALMA images. The simulated and original ALMA images have the same size, therefore the impact of simultaneous multi-source fitting is also considered. Moreover, for the prior fitting with GALFIT, we use the catalog of simulated galaxies as the prior source list.

Similar to Fig. C.1, we present the comparison of the simulated and recovered fluxes from the "PHYS" simulation in Fig. C.2. The "PHYS" simulation contains many more faint sources (due to the realistic stellar mass function and main sequence correlation). The histograms in the right panels of Fig. C.2 are narrower than those of the "FULL" simulation as shown in Fig. C.1, especially for the low S/N_{peak} sources (shown as the red histograms in both figures). This means that the flux bias and error are different if we adopt different simulation methods (Fig. 12 and 13), even when each simulation is repeated sufficiently to yield robust statistics. Here we emphasize that the prior information assumed in the simulations is important for analyzing the flux bias and error statistics, and making the simulation as close to real galaxy population distribution as possible will lead to more realistic results (Sect. 3.1).

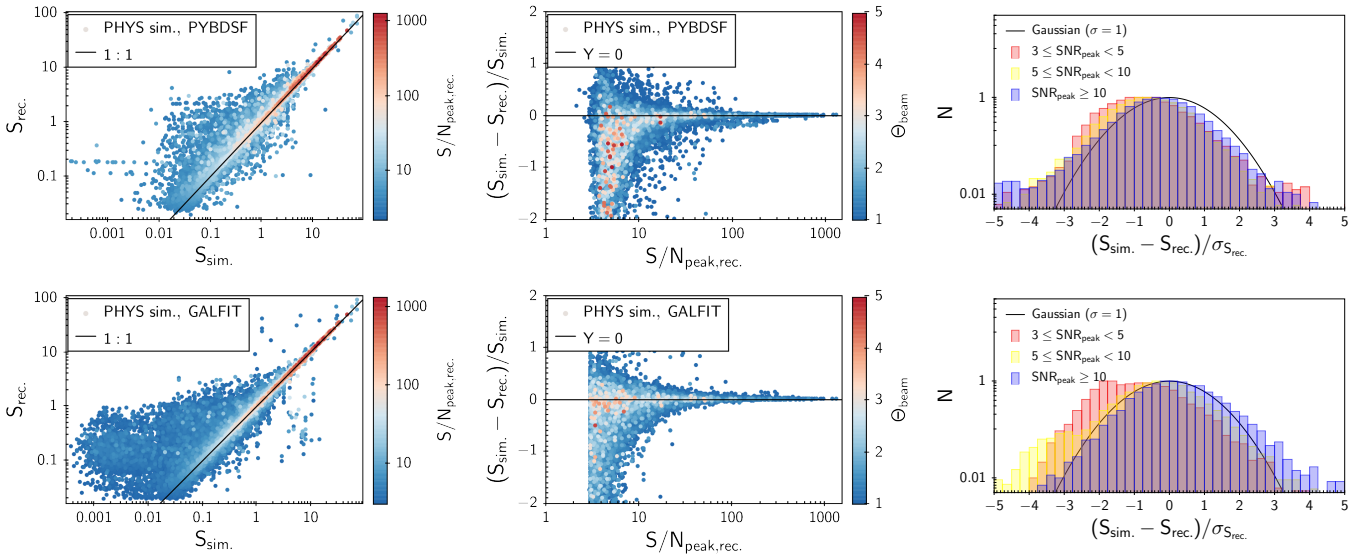

Fig. C.2. Similar to Fig. C.1 but for "PHYS" simulation. See Fig. C.1 caption.

C.2.4. *Limitations*

The physically-motivated simulation has the advantage of resembling closer the real situation for galaxy photometry in ALMA images. However, it also has limitations, both in the assumed galaxy evolution models and when comparing to the full-parameter-space simulation.

Firstly, our model galaxies are built on the star-forming galaxy's stellar mass function at each redshift. These mass functions are not well constrained at redshift of $z \sim 4$ and unconstrained at higher redshifts (e.g., Grazian et al. 2015; Song et al. 2016; Davidzon et al. 2017). Then, we assume a starburst fraction associated to the merger fraction, which is highly unconstrained at redshift $z \sim 2$ and beyond. Moreover, because we aim to reproduce the majority of star-forming galaxies, we choose simplified galaxy SED models (Magdis et al. 2012; Liu et al. 2018) which can represent the bulk of star-forming and starburst galaxies (see details in Appx. C.2.1 and references there), but these SED models do not include extreme cases, e.g., galaxies with very high sSFR, very low or very high dust temperature, etc. The "PHYS" simulation also does not include a full implementation of the clustering effect as in Béthermin et al. (2017). However, we emphasize that the aim of the "PHYS" simulation is to provide very different inputs from the "FULL" simulation, to see whether they can lead to different statistical results, and they do. Further the "PHYS" simulation we adopt here is sufficiently complex for testing our ALMA (sub-)mm photometry under all possible, physical situations that are not covered by the "FULL" simulation.

C.3. *Final statistical behaviour of corrected fluxes and errors*

We provide details on the statistical behaviour of flux errors here as an extension to the discussion in Appx. 3.1.5. We examined the histograms of $(S_{\rm sim.} - S^{\rm corr.}_{\rm rec.})/\sigma_{S^{\rm corr.}_{\rm rec.}}$ as shown in Fig. C.3 for final corrected PyBDSF and GALFIT fluxes and errors. Such a histogram indicates how well our final flux errors $\sigma_{S^{\rm corr.}_{\rm rec.}}$ can reflect the true scatter between $S_{\rm sim.}$ and $S^{\rm corr.}_{\rm rec.}$. Ideally, if the flux error well represents the uncertainty in the photometry, the histogram should have the shape of an 1D Gaussian with mean = 0 and sigma = 1 (which is overlaid in the figure). Comparing these histograms to those before correction (Fig. C.1 and C.2), we do find significant improvement in the shape of the histogram. In Fig. C.3, we show the histogram after each step of correction: (1) flux bias correction (*top row*) and (2) both flux bias and error correction (*bottom row*). The final histograms nicely agree with the mean = 0, sigma = 1 1D Gaussian. Although there are some outliers, but they only contribute a few percent in number. The outlier fraction is lower in the GALFIT photometry compared to the PyBDSF photometry, probably because prior fitting uses known positional information and reduces the chance of recovering noise peaks as sources. Both of these effects are also related to the features in each photometry method. For example, at the high value-end of the histogram, outliers have highly underestimated fluxes likely due to the aforementioned resolution bias (large sources are break down and only partial flux are recovered). And at the low-value end, outliers have extra-boosted fluxes mostly because the recovered source

sizes are significantly larger than their simulated sizes. For these outliers, we can tentatively identify them out by checking their PyBDSF multi-component flag (`Flag_multi`), and also comparing their PyBDSF and GALFIT fluxes and sizes. For example, there are 5% sources with `Flag_multi=='M'` in our blind photometry catalog. Therefore, if excluding them, for the bulk of sources done with out photometry, the errors are quite well behaved in statistics.

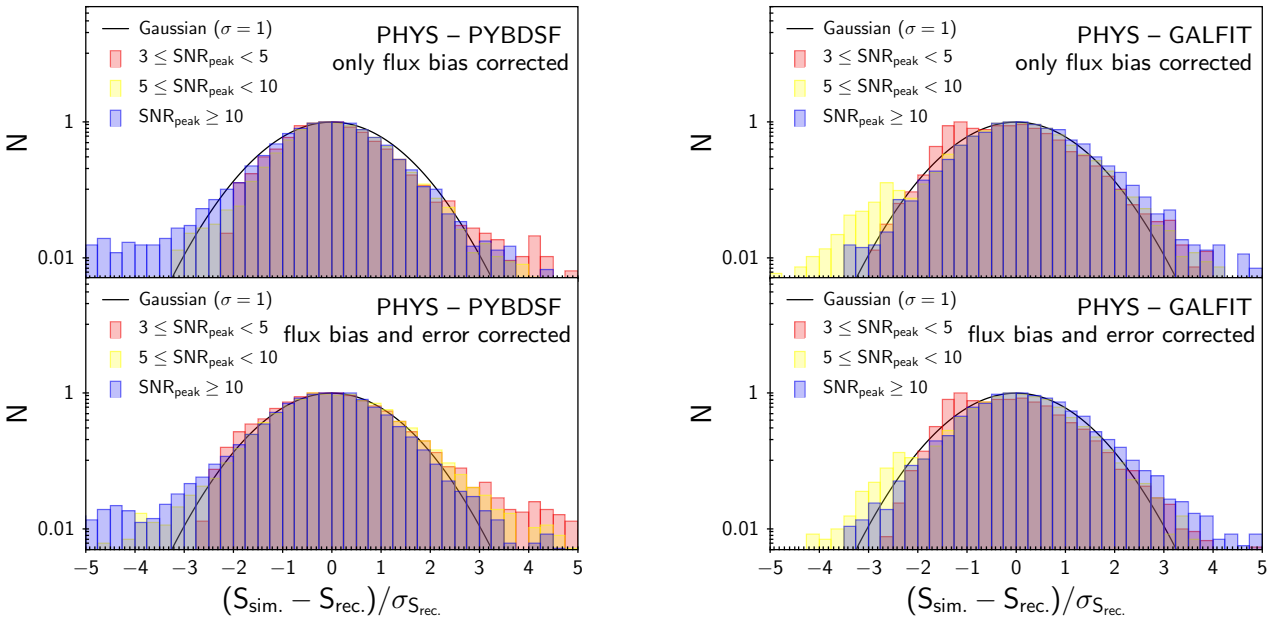

Fig. C.3. Left panels are based on the "PHYS" simulation with PyBDSF photometry, and **right panels** are based on the "PHYS" simulation with GALFIT photometry. In each left/right panel, the *upper panel* is the histogram of $(S_{\rm sim.} - S^{\rm corr.}_{\rm rec.})/\sigma_{S_{\rm rec.},\,\rm Condon1997}$, where $S^{\rm corr.}_{\rm rec.}$ is the measured total flux after flux bias correction (Sect. 3.1.1), and $\sigma_{S_{\rm rec.},\,\rm Condon1997}$ is the error in total flux shipped with PyBDSF based on Condon (1997). And the *lower panel* is the histogram of $(S_{\rm sim.} - S^{\rm corr.}_{\rm rec.})/\sigma_{S^{\rm corr.}_{\rm rec.}}$, where $\sigma_{S^{\rm corr.}_{\rm rec.}}$ is the simulation-based error in total flux (Sect. 3.1.3). An 1D Gaussian with mean = 0 and sigma = 1 is overlaid in each panel.

C.4. *Discussion on the completeness of (sub-)mm/radio photometry*

We present the 2D diagnostic diagram of completeness in Fig. C.4. Such a diagram is also used in similar works (e.g., Jiménez-Andrade et al. 2019; Franco et al. 2018). The left panel shows the dependency on $S_{\rm peak}$ and the right panel on $S_{\rm total}$. Because $S_{\rm total} \propto S_{\rm peak} \times \Theta^2_{\rm beam,\,convol.}$, the completeness has a more complicated dependency on $S_{\rm total}$/rms noise than $\mathrm{S/N_{peak}}$. Thus using $\mathrm{S/N_{peak}}$ to select the sample results in a more uniform completeness for various source sizes.

We compare our findings with other ALMA photometry work with completeness assessments in the literature: Hatsukade et al. (2011, 2016), Karim et al. (2013), Ono et al. (2014), Simpson et al. (2015), Aravena et al. (2016), Umehata et al. (2017), Dunlop et al. (2017) and Franco et al. (2018). Our differential completeness at low $\mathrm{S/N_{peak}}$ is either lower than or consistent with the values in the literature. For example, Karim et al. (2013) estimated ∼70% completeness at $\mathrm{S/N_{peak}} \sim 3$ for their Gaussian-fitting photometry down to $2.5\,\sigma$ while we derive a value of ∼20%. Aravena et al. (2016) estimated ∼50% completeness at $\mathrm{S/N_{peak}} \sim 3$ for their SExtractor (Bertin & Arnouts 1996) source extraction down to $3.5\,\sigma$. Hatsukade et al. (2016) found ∼25% completeness at $\mathrm{S/N_{peak}} \sim 3$ for their AEGEAN (Hancock et al. 2012) Gaussian-fitting photometry down to $4\,\sigma$. And Franco et al. (2018) report <20% completeness at the same $\mathrm{S/N_{peak}}$ for their Blobcat Gaussian-fitting photometry, which is consistent with ours. Note that all these studies mentioned are based on ALMA maps coming from a single program with similar beam size and rms noise. However, in our work, these parameters vary significantly across the archival programs. Based on archival data, Ono et al. (2014) found a completeness of ∼85% at $\mathrm{S/N_{peak}} \sim 3$ for their SExtractor extraction down to as low as $1.8\,\sigma$.

Our comparison above shows that both detection criterion and photometry method are important when deriving completeness fractions. A lower detection criterion leads to a higher completeness. However, we have also to consider

the spurious detection fraction (Fig. 8) which rapidly increases by lowering the detection criterion. Our choice of PyBDSF parameters as described in Sect. 2.2 are thus a compromise between the completeness and spurious fractions.

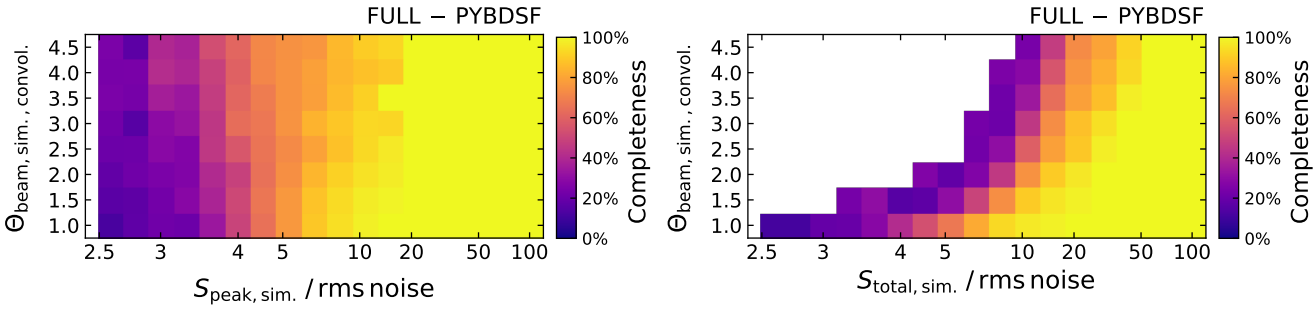

Fig. C.4. Completeness of the PyBDSF source extraction as a 2D function of the simulated $S_{\rm peak}$/rms noise (i.e., S/N$_{\rm peak}$; left panel) or $S_{\rm total}$/rms noise (right panel) and $\Theta_{\rm beam}$, based on the "FULL" simulation (Sect. C.1). Color indicates the completeness fraction.

D. AN EXAMPLE OF OUR AUTOMATED COUNTERPART ASSOCIATION EXAMINATION

In Fig. D.1 we show an example for our automated examination of the counterpart association between each ALMA source and its counterpart source in prior catalogs. For each counterpart image (HST ACS i-band, UltraVISTA K_s-band, SPLASH IRAC ch1, and VLA 3 GHz), we measure several parameters as described in Sect. 4.2, then link them to the likelihood of the counterpart association based on our visual inspection. For example, in Fig. D.1, the bold green circle indicates the fitted ALMA source position and size (convolved with the ALMA beam), and the red crosses are sources in the prior catalog (Sect. 2.3). The bold red cross indicates the counterpart of the ALMA source (or just the prior source used in prior fitting photometry).

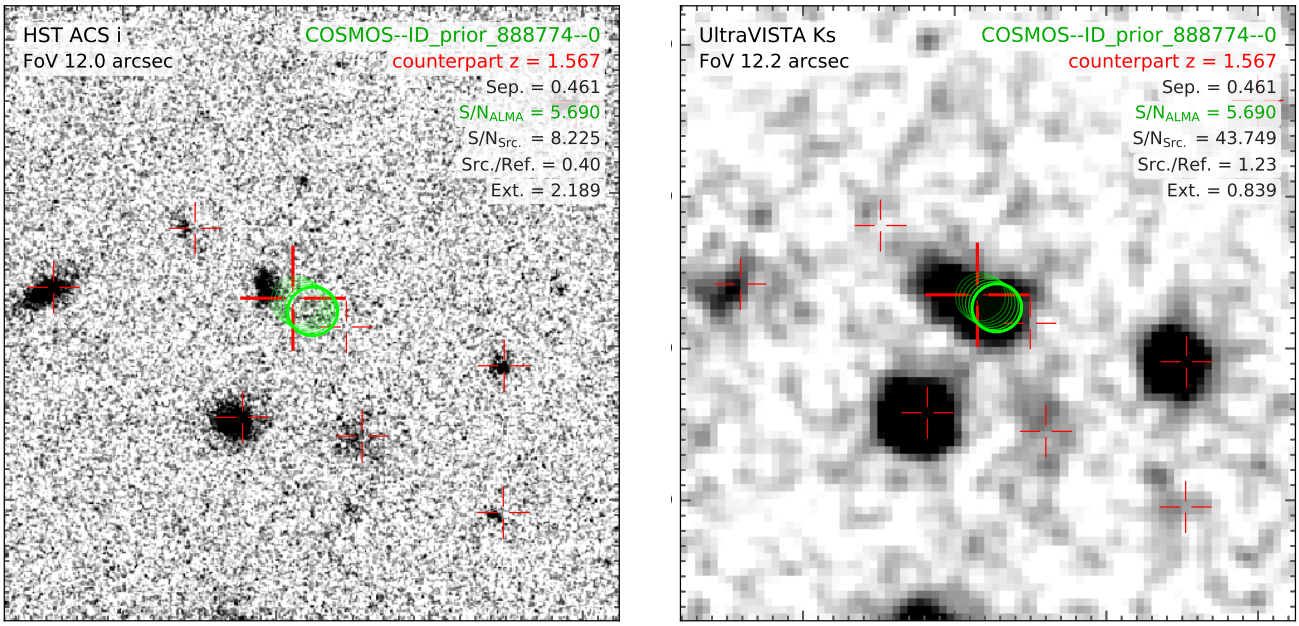

Fig. D.1. Example of our automated counterpart association (Sect. 4.2) for the ALMA source with COSMOS2015 ID 888774 (Laigle et al. 2016). It is selected as an example because of its relatively large offset (Sep. = 0.46; see definition in Sect. 4.2) between the ALMA and optical positions (corrected for astrometry). The background image is the HST ACS i-band in the **left** panel and the UltraVISTA K_s-band in the **right** panel. Other symbols are the same in two panels: the thick green ellipse shows the position and size of the ALMA source, and the thin green ellipses show the aperture used for measuring the **Ext**. parameter as described in Sect. 4.2. The largest red cross represents the counterpart position and smaller red crosses are other master catalog sources within this image. This prior source is from the Laigle et al. (2016) catalog, therefore its prior position is based on an optical-to-near-IR combined SNR map and centered at the K_s emission. There is a faint blue source with ID 1732293 in the Capak et al. (2007) i-band catalog (but not in the Laigle et al. (2016) catalog) at the North-East of the prior position with a small offset of 0.4″. Current information (including astrometry) is not sufficient to distinguish whether this faint blue source is another galaxy or physically associated to the K_s source. The latter situation (where UV stellar light offsets from dust emission) has been observed in many high-redshift dusty galaxies (e.g., Hodge et al. 2016, 2019; Rujopakarn et al. 2019; Lang et al. 2019). Due to the fact that the i-band source is quite faint, the K_s galaxy's SED is mostly unaffected by the i-band and shorter wavelength photometry, and the ALMA photometry can be reasonably fitted for the K_s galaxy at its redshift.

E. PROBLEMATIC SED FITTING CASES

In Fig. E.1 we show the four problematic SED fitting cases which are labeled in Fig. 24 and discussed in Sect. 4.6. Their stellar masses derived from our optimized iterative MAGPHYS fitting (Sect. 4.4) without an AGN component are higher by a factor of 10–100 compared to the masses reported by Delvecchio et al. (2017), who used SED3FIT to account for an AGN SED component (and both redshifts are consistent). However, we have following reasons to believe that these are just rare cases which do not indicate an obvious bias in our SED fitting affected by mid-IR AGN contamination. Firstly, these sources are very rare, only four out of a total of 396 sources cross-matched with the Delvecchio et al. (2017) catalog. All other AGNs have no such large difference between this work and Delvecchio et al. (2017). Secondly, the advantage of this work is that we have ALMA and far-IR/(sub-)mm constraint, and MAGPHYS has the energy balance assumption which links optical dust attenuation to dust luminosity in IR, therefore the degeneracy between age and attenuation can be reduced (if the energy balance is valid, which could be true for our ALMA-selected dusty sample). Thirdly, the optical photometry itself has uncertainties from either noise or galaxy-galaxy blending. E.g., for outlier 3, its HST i-band and Subaru Suprime-Cam z'-band data could not be fit well, but its IR SED looks reasonable and counterpart association shows no problem. It has no obvious blended optical source within 3″, but we could not rule out the chance of line-of-sight blending of two sources, i.e., like the case of the galaxy CRLE reported in Pavesi et al. 2018.

To verify the degeneracy between age and attenuation, we ran some additional SED fitting with our multi-component χ^2 fitting code under development [27], similar to Liu et al. (2018). We fit two components for test purposes: one uses Bruzual & Charlot (2003) stellar SED models at solar metallicity, with constant star formation history, with various ages from 0.1 to 1 Gyr and with a varied dust attenuation ($E(B-V) = 0$ to 1.2) following the Calzetti et al. (2000) attenuation law; and the other uses the Mullaney et al. (2011) AGN SED models. We fit only the optical to mid-IR part ($\lambda < 20\,\mu m$) of the SED. We loop each combination of the two component models and find a best χ^2 fit, then merge all combinations together to analyze the χ^2 distributions for M_* and age (to obtain a median value and error for each parameter, similar to Liu et al. 2018). We find that such a fitting without constraints from far-IR dust emission leads to very large uncertainties in age and stellar mass. For example, for outlier `2` (ID 951838), the best-fit $\log_{10}(M_*/\mathrm{M}_\odot)$ dramatically varies from 9.5 ± 0.5 to 11.4 ± 0.6 with an age varied from 200 Myr to 1 Gyr (with best-fit $E(B-V) = 0.5$ and 0.7 respectively). For comparison, the Delvecchio et al. (2017) $\log_{10} M_*$ of 10.24 and our MAGPHYS value of 12.15 agreed within the errors. For our outlier `3` (ID 813955), a fixed age of 200 Myr fitting gives $E(B-V) = 0.5$ and $\log_{10} M_* = 11.0 \pm 0.1$, while a free-age fitting gives $E(B-V) = 1.2$ and $\log_{10} M_* = 12.6 \pm 0.3$ with an age of 450 Myr. In comparison, Delvecchio et al. (2017) report a $\log_{10} M_*$ of 10.34 and our MAGPHYS value is 11.63, also agreed within the range of ucertainties. Outlier `4` (ID 842140) presents a similar situation. While for outlier `1` (ID 422662), our experimental stellar+AGN SED fitting ($\log_{10} M_* \sim 11.2 - 12.6 \pm 0.6$) cannot not account for the low $\log_{10} M_* = 9.45$ reported in the Delvecchio et al. (2017) catalog (their `ID_VLA3` 4136). Therefore, as mentioned in the text of Sect. 4.6, we think its stellar mass is more reliable from the MAGPHYS fitting because of the inclusion of ALMA data here. Finally, we conclude that these sources are just low probability outliers suffering from the large uncertainty in stellar mass estimation and possibly also optical line-of-sight blending. The latter needs further follow-up observations which is beyond the scope of this paper.

[27] It is still in the experimental phase but is available at http://github.com/1054/Crab.Toolkit.michi2.

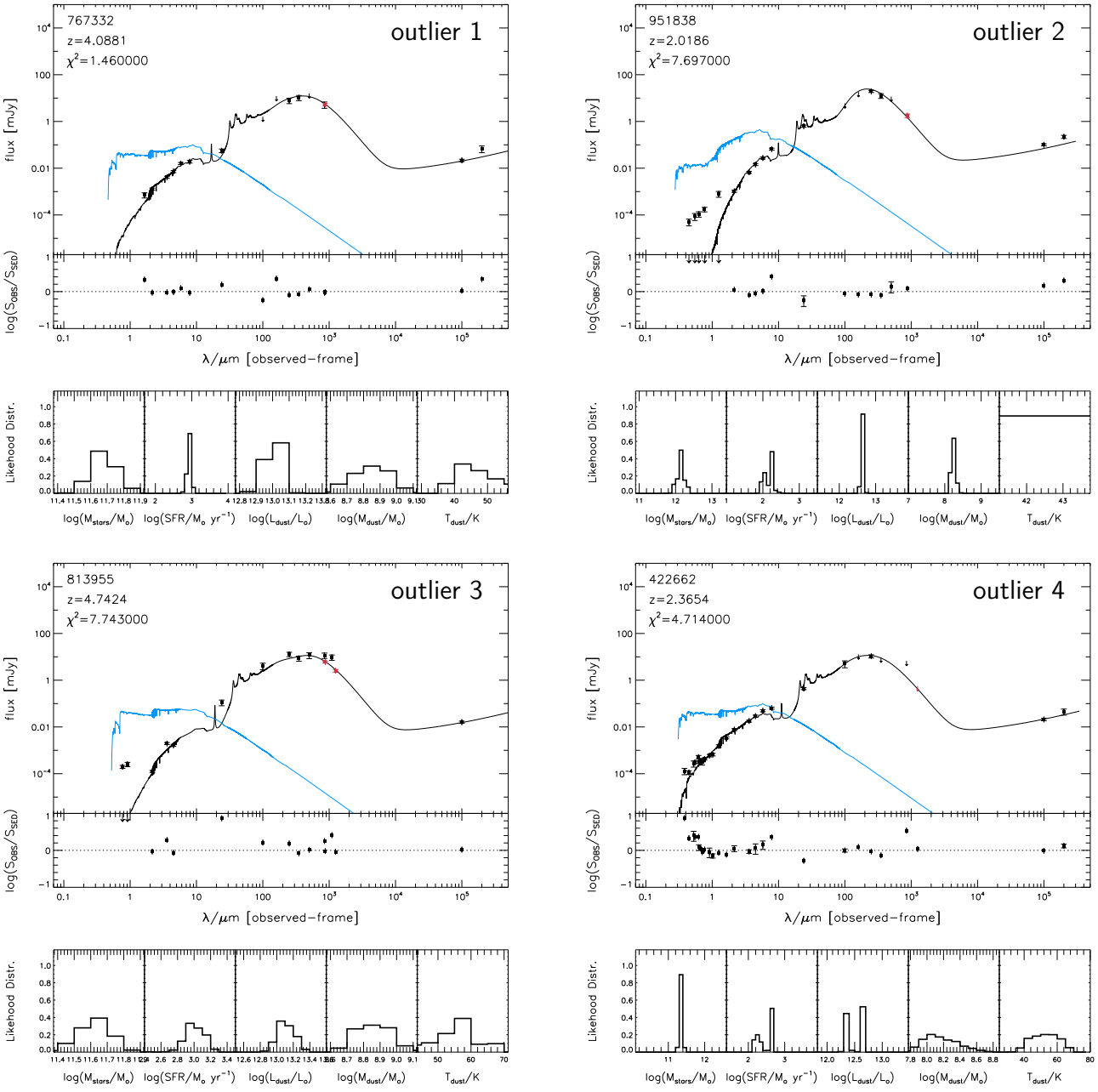

Fig. E.1. Example of SEDs for sources 1, 2, 3 and 4 listed in Fig. 24 which exhibit strong mid-IR AGN emission and disagreed in the stellar masses between this work and Delvecchio et al. (2017). See discussion of each source in Sect. 4.6. For each source, the best-fit full SED (black line) and unattenuated stellar SED (cyan line) as well as the photometric data points are shown in the upper, large sub-panel. The ALMA data point from this work is highlighted in red. Optical to near-IR K band data are from the "COSMOS2015" catalog (Laigle et al. 2016) with 3″ aperture photometry; near-IR IRAC to mm and radio data are from the "super-deblended" FIR/mm catalog (Jin et al. 2018). The SED fitting residuals at each band ($\log_{10} S_{\rm OBS}/S_{\rm SED}$) are shown in each middle sub-panel. The probability distribution histograms of each fitted physical parameter ($\log_{10} M_*$, $\log_{10}$ SFR, IR luminosity $\log_{10} L_{\rm dust}$, dust mass $\log_{10} M_{\rm dust}$ and dust temperature $T_{\rm dust}$) are shown in the sub-panels at the bottom.

REFERENCES

Albrecht, M., Krügel, E., & Chini, R. 2007, A&A, 462, 575, doi: 10.1051/0004-6361:20047017

Aravena, M., Decarli, R., Walter, F., et al. 2016, ApJ, 833, 68, doi: 10.3847/1538-4357/833/1/68

Aretxaga, I., Wilson, G. W., Aguilar, E., et al. 2011, MNRAS, 415, 3831, doi: 10.1111/j.1365-2966.2011.18989.x

Ashby, M. L. N., Caputi, K. I., Cowley, W., et al. 2018, ApJS, 237, 39, doi: 10.3847/1538-4365/aad4fb

Baan, W. A., Henkel, C., Loenen, A. F., Baudry, A., & Wiklind, T. 2008, A&A, 477, 747, doi: 10.1051/0004-6361:20077203

Bauermeister, A., Blitz, L., Bolatto, A., et al. 2013, ApJ, 763, 64, doi: 10.1088/0004-637X/763/1/64

Berta, S., Lutz, D., Santini, P., et al. 2013, A&A, 551, A100, doi: 10.1051/0004-6361/201220859

Bertemes, C., Wuyts, S., Lutz, D., et al. 2018, MNRAS, 478, 1442, doi: 10.1093/mnras/sty963

Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393, doi: 10.1051/aas:1996164

Bertoldi, F., Carilli, C., Aravena, M., et al. 2007, ApJS, 172, 132, doi: 10.1086/520511

Béthermin, M., Daddi, E., Magdis, G., et al. 2012a, ApJL, 757, L23, doi: 10.1088/2041-8205/757/2/L23

Béthermin, M., Le Floc'h, E., Ilbert, O., et al. 2012b, A&A, 542, A58, doi: 10.1051/0004-6361/201118698

Béthermin, M., Daddi, E., Magdis, G., et al. 2015, A&A, 573, A113, doi: 10.1051/0004-6361/201425031

Béthermin, M., Wu, H.-Y., Lagache, G., et al. 2017, A&A, 607, A89, doi: 10.1051/0004-6361/201730866

Bigiel, F., Leroy, A., Walter, F., et al. 2008, AJ, 136, 2846, doi: 10.1088/0004-6256/136/6/2846

Blain, A. W., Smail, I., Ivison, R. J., Kneib, J.-P., & Frayer, D. T. 2002, PhR, 369, 111, doi: 10.1016/S0370-1573(02)00134-5

Bolatto, A. D., Wolfire, M., & Leroy, A. K. 2013, ARA&A, 51, 207, doi: 10.1146/annurev-astro-082812-140944

Bondi, M., Ciliegi, P., Schinnerer, E., et al. 2008, ApJ, 681, 1129, doi: 10.1086/589324

Bondi, M., Ciliegi, P., Zamorani, G., et al. 2003, A&A, 403, 857, doi: 10.1051/0004-6361:20030382

Brinchmann, J., Charlot, S., White, S. D. M., et al. 2004, MNRAS, 351, 1151, doi: 10.1111/j.1365-2966.2004.07881.x

Brusa, M., Feruglio, C., Cresci, G., et al. 2015, A&A, 578, A11, doi: 10.1051/0004-6361/201425491

Brusa, M., Cresci, G., Daddi, E., et al. 2018, A&A, 612, A29, doi: 10.1051/0004-6361/201731641

Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000, doi: 10.1046/j.1365-8711.2003.06897.x

Bussmann, R. S., Pérez-Fournon, I., Amber, S., et al. 2013, ApJ, 779, 25, doi: 10.1088/0004-637X/779/1/25

Bussmann, R. S., Riechers, D., Fialkov, A., et al. 2015, ApJ, 812, 43, doi: 10.1088/0004-637X/812/1/43

Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, ApJ, 533, 682, doi: 10.1086/308692

Capak, P., Aussel, H., Ajiki, M., et al. 2007, ApJS, 172, 99, doi: 10.1086/519081

Capak, P. L., Carilli, C., Jones, G., et al. 2015, Nature, 522, 455, doi: 10.1038/nature14500

Carilli, C. L., & Walter, F. 2013, ARA&A, 51, 105, doi: 10.1146/annurev-astro-082812-140953

Carniani, S., Maiolino, R., De Zotti, G., et al. 2015, A&A, 584, A78, doi: 10.1051/0004-6361/201525780

Carpenter, J., Iono, D., Testi, L., et al. 2019, arXiv e-prints. https://arxiv.org/abs/1902.02856

Casey, C. M., Narayanan, D., & Cooray, A. 2014, PhR, 541, 45, doi: 10.1016/j.physrep.2014.02.009

Casey, C. M., Berta, S., Béthermin, M., et al. 2012, ApJ, 761, 140, doi: 10.1088/0004-637X/761/2/140

Casey, C. M., Cooray, A., Killi, M., et al. 2017, ApJ, 840, 101, doi: 10.3847/1538-4357/aa6cb1

Chabrier, G. 2003, PASP, 115, 763, doi: 10.1086/376392

Chang, Y.-Y., van der Wel, A., da Cunha, E., & Rix, H.-W. 2015, ApJS, 219, 8, doi: 10.1088/0067-0049/219/1/8

Chen, C.-C., Hodge, J. A., Smail, I., et al. 2017, ApJ, 846, 108, doi: 10.3847/1538-4357/aa863a

Civano, F., Elvis, M., Brusa, M., et al. 2012, ApJS, 201, 30, doi: 10.1088/0067-0049/201/2/30

Civano, F., Marchesi, S., Comastri, A., et al. 2016, ApJ, 819, 62, doi: 10.3847/0004-637X/819/1/62

Comparat, J., Richard, J., Kneib, J.-P., et al. 2015, A&A, 575, A40, doi: 10.1051/0004-6361/201424767

Condon, J. J. 1997, PASP, 109, 166, doi: 10.1086/133871

Conselice, C. J. 2014, ARA&A, 52, 291, doi: 10.1146/annurev-astro-081913-040037

Coogan, R. T., Sargent, M. T., Daddi, E., et al. 2019, MNRAS, 485, 2092, doi: 10.1093/mnras/stz409

Cooke, E. A., Smail, I., Swinbank, A. M., et al. 2018, ApJ, 861, 100, doi: 10.3847/1538-4357/aac6ba

Cresci, G., Mainieri, V., Brusa, M., et al. 2015, ApJ, 799, 82, doi: 10.1088/0004-637X/799/1/82

da Cunha, E., Charlot, S., & Elbaz, D. 2008, MNRAS, 388, 1595, doi: 10.1111/j.1365-2966.2008.13535.x

da Cunha, E., Walter, F., Smail, I. R., et al. 2015, ApJ, 806, 110, doi: 10.1088/0004-637X/806/1/110

Daddi, E., Dickinson, M., Morrison, G., et al. 2007, ApJ, 670, 156, doi: 10.1086/521818
Daddi, E., Dannerbauer, H., Liu, D., et al. 2015, A&A, 577, A46, doi: 10.1051/0004-6361/201425043
Davidzon, I., Ilbert, O., Laigle, C., et al. 2017, A&A, 605, A70, doi: 10.1051/0004-6361/201730419
De Looze, I., Baes, M., Bendo, G. J., Cortese, L., & Fritz, J. 2011, MNRAS, 416, 2712, doi: 10.1111/j.1365-2966.2011.19223.x
Delvecchio, I., Smolčić, V., Zamorani, G., et al. 2017, A&A, 602, A3, doi: 10.1051/0004-6361/201629367
Draine, B. T., & Li, A. 2007, ApJ, 657, 810, doi: 10.1086/511055
Dunlop, J. S., McLure, R. J., Biggs, A. D., et al. 2017, MNRAS, 466, 861, doi: 10.1093/mnras/stw3088
Elbaz, D., Daddi, E., Le Borgne, D., et al. 2007, A&A, 468, 33, doi: 10.1051/0004-6361:20077525
Elbaz, D., Dickinson, M., Hwang, H. S., et al. 2011, A&A, 533, A119, doi: 10.1051/0004-6361/201117239
Elvis, M., Civano, F., Vignali, C., et al. 2009, ApJS, 184, 158, doi: 10.1088/0067-0049/184/1/158
Franco, M., Elbaz, D., Béthermin, M., et al. 2018, A&A, 620, A152, doi: 10.1051/0004-6361/201832928
Fu, H., Yan, L., Scoville, N. Z., et al. 2010, ApJ, 722, 653, doi: 10.1088/0004-637X/722/1/653
Fujimoto, S., Ouchi, M., Shibuya, T., & Nagai, H. 2017, ApJ, 850, 83, doi: 10.3847/1538-4357/aa93e6
Geach, J. E., Dunlop, J. S., Halpern, M., et al. 2017, MNRAS, 465, 1789, doi: 10.1093/mnras/stw2721
Ginsburg, A., Sipőcz, B. M., Brasseur, C. E., et al. 2019, AJ, 157, 98, doi: 10.3847/1538-3881/aafc33
Grazian, A., Fontana, A., Santini, P., et al. 2015, A&A, 575, A96, doi: 10.1051/0004-6361/201424750
Greisen, E. W. 2003, in Astrophysics and Space Science Library, Vol. 285, Information Handling in Astronomy - Historical Vistas, ed. A. Heck, 109
Groves, B. A., Schinnerer, E., Leroy, A., et al. 2015, ApJ, 799, 96, doi: 10.1088/0004-637X/799/1/96
Hales, C. A., Murphy, T., Curran, J. R., et al. 2012, MNRAS, 425, 979, doi: 10.1111/j.1365-2966.2012.21373.x
Hancock, P. J., Murphy, T., Gaensler, B. M., Hopkins, A., & Curran, J. R. 2012, MNRAS, 422, 1812, doi: 10.1111/j.1365-2966.2012.20768.x
Hasinger, G., Capak, P., Salvato, M., et al. 2018, ApJ, 858, 77, doi: 10.3847/1538-4357/aabacf
Hatsukade, B., Kohno, K., Aretxaga, I., et al. 2011, MNRAS, 411, 102, doi: 10.1111/j.1365-2966.2010.17658.x
Hatsukade, B., Kohno, K., Umehata, H., et al. 2016, PASJ, 68, 36, doi: 10.1093/pasj/psw026
Hatsukade, B., Kohno, K., Yamaguchi, Y., et al. 2018, PASJ, 70, 105, doi: 10.1093/pasj/psy104
Hayward, C. C., & Smith, D. J. B. 2015, MNRAS, 446, 1512, doi: 10.1093/mnras/stu2195
Hezaveh, Y. D., & Holder, G. P. 2011, ApJ, 734, 52, doi: 10.1088/0004-637X/734/1/52
Hilbert, S., White, S. D. M., Hartlap, J., & Schneider, P. 2007, MNRAS, 382, 121, doi: 10.1111/j.1365-2966.2007.12391.x
Hodge, J. A., Swinbank, A. M., Simpson, J. M., et al. 2016, ApJ, 833, 103, doi: 10.3847/1538-4357/833/1/103
Hodge, J. A., Smail, I., Walter, F., et al. 2019, ApJ, 876, 130, doi: 10.3847/1538-4357/ab1846
Hughes, T. M., Ibar, E., Villanueva, V., et al. 2017, MNRAS, 468, L103, doi: 10.1093/mnrasl/slx033
Hunt, L. K., De Looze, I., Boquien, M., et al. 2019, A&A, 621, A51, doi: 10.1051/0004-6361/201834212
Ilbert, O., McCracken, H. J., Le Fèvre, O., et al. 2013, A&A, 556, A55, doi: 10.1051/0004-6361/201321100
Jiménez-Andrade, E. F., Magnelli, B., Karim, A., et al. 2019, A&A, 625, A114, doi: 10.1051/0004-6361/201935178
Jin, S., Daddi, E., Liu, D., et al. 2018, ApJ, 864, 56, doi: 10.3847/1538-4357/aad4af
Karim, A., Swinbank, A. M., Hodge, J. A., et al. 2013, MNRAS, 432, 2, doi: 10.1093/mnras/stt196
Kennicutt, Jr., R. C. 1998, ApJ, 498, 541, doi: 10.1086/305588
Kriek, M., Shapley, A. E., Reddy, N. A., et al. 2015, ApJS, 218, 15, doi: 10.1088/0067-0049/218/2/15
Laigle, C., McCracken, H. J., Ilbert, O., et al. 2016, ApJS, 224, 24, doi: 10.3847/0067-0049/224/2/24
Lang, P., Schinnerer, E., Smail, I., et al. 2019, arXiv e-prints, arXiv:1905.06960. https://arxiv.org/abs/1905.06960
Le Fèvre, O., Tasca, L. A. M., Cassata, P., et al. 2015, A&A, 576, A79, doi: 10.1051/0004-6361/201423829
Le Floc'h, E., Aussel, H., Ilbert, O., et al. 2009, ApJ, 703, 222, doi: 10.1088/0004-637X/703/1/222
Leauthaud, A., Massey, R., Kneib, J.-P., et al. 2007, ApJS, 172, 219, doi: 10.1086/516598
Lee, N., Sheth, K., Scott, K. S., et al. 2017, MNRAS, 471, 2124, doi: 10.1093/mnras/stx1753
Leitherer, C., & Heckman, T. M. 1995, ApJS, 96, 9, doi: 10.1086/192112
Leroy, A. K., Walter, F., Brinks, E., et al. 2008, AJ, 136, 2782, doi: 10.1088/0004-6256/136/6/2782
Leroy, A. K., Bolatto, A., Gordon, K., et al. 2011, ApJ, 737, 12, doi: 10.1088/0004-637X/737/1/12

Lilly, S. J., Le Fèvre, O., Renzini, A., et al. 2007, ApJS, 172, 70, doi: 10.1086/516589
Lilly, S. J., Le Brun, V., Maier, C., et al. 2009, ApJS, 184, 218, doi: 10.1088/0067-0049/184/2/218
Liu, D., Gao, Y., Isaak, K., et al. 2015, ApJL, 810, L14, doi: 10.1088/2041-8205/810/2/L14
Liu, D., Daddi, E., Dickinson, M., et al. 2018, ApJ, 853, 172, doi: 10.3847/1538-4357/aaa600
Lutz, D., Poglitsch, A., Altieri, B., et al. 2011, A&A, 532, A90, doi: 10.1051/0004-6361/201117107
Madau, P., & Dickinson, M. 2014, ARA&A, 52, 415, doi: 10.1146/annurev-astro-081811-125615
Magdis, G. E., Daddi, E., Elbaz, D., et al. 2011, ApJL, 740, L15, doi: 10.1088/2041-8205/740/1/L15
Magdis, G. E., Daddi, E., Béthermin, M., et al. 2012, ApJ, 760, 6, doi: 10.1088/0004-637X/760/1/6
Magdis, G. E., Rigopoulou, D., Daddi, E., et al. 2017, A&A, 603, A93, doi: 10.1051/0004-6361/201731037
Magnelli, B., Saintonge, A., Lutz, D., et al. 2012, A&A, 548, A22, doi: 10.1051/0004-6361/201220074
Marchesi, S., Civano, F., Elvis, M., et al. 2016, ApJ, 817, 34, doi: 10.3847/0004-637X/817/1/34
Marsan, Z. C., Marchesini, D., Brammer, G. B., et al. 2017, ApJ, 842, 21, doi: 10.3847/1538-4357/aa7206
Masters, D. C., Stern, D. K., Cohen, J. G., et al. 2017, ApJ, 841, 111, doi: 10.3847/1538-4357/aa6f08
McCracken, H. J., Capak, P., Salvato, M., et al. 2010, ApJ, 708, 202, doi: 10.1088/0004-637X/708/1/202
McCracken, H. J., Milvang-Jensen, B., Dunlop, J., et al. 2012, A&A, 544, A156, doi: 10.1051/0004-6361/201219507
McMullin, J. P., Waters, B., Schiebel, D., Young, W., & Golap, K. 2007, in Astronomical Society of the Pacific Conference Series, Vol. 376, Astronomical Data Analysis Software and Systems XVI, ed. R. A. Shaw, F. Hill, & D. J. Bell, 127
Miettinen, O., Delvecchio, I., Smolčić, V., et al. 2017a, A&A, 597, A5, doi: 10.1051/0004-6361/201628128
—. 2017b, A&A, 606, A17, doi: 10.1051/0004-6361/201730762
Mohan, N., & Rafferty, D. 2015, PyBDSF: Python Blob Detection and Source Finder, Astrophysics Source Code Library. http://ascl.net/1502.007
Mullaney, J. R., Alexander, D. M., Goulding, A. D., & Hickox, R. C. 2011, MNRAS, 414, 1082, doi: 10.1111/j.1365-2966.2011.18448.x
Muzzin, A., Marchesini, D., Stefanon, M., et al. 2013, ApJ, 777, 18, doi: 10.1088/0004-637X/777/1/18
Nanayakkara, T., Glazebrook, K., Kacprzak, G. G., et al. 2016, ApJ, 828, 21, doi: 10.3847/0004-637X/828/1/21
Negrello, M., Hopwood, R., De Zotti, G., et al. 2010, Science, 330, 800, doi: 10.1126/science.1193420
Noeske, K. G., Weiner, B. J., Faber, S. M., et al. 2007, ApJL, 660, L43, doi: 10.1086/517926
Oliver, S. J., Bock, J., Altieri, B., et al. 2012, MNRAS, 424, 1614, doi: 10.1111/j.1365-2966.2012.20912.x
Ono, Y., Ouchi, M., Kurono, Y., & Momose, R. 2014, ApJ, 795, 5, doi: 10.1088/0004-637X/795/1/5
Pavesi, R., Riechers, D. A., Sharon, C. E., et al. 2018, ApJ, 861, 43, doi: 10.3847/1538-4357/aac6b6
Pearson, W. J., Wang, L., Hurley, P. D., et al. 2018, A&A, 615, A146, doi: 10.1051/0004-6361/201832821
Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, Journal of Machine Learning Research, 12, 2825
Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2002, AJ, 124, 266, doi: 10.1086/340952
—. 2010, AJ, 139, 2097, doi: 10.1088/0004-6256/139/6/2097
Perna, M., Brusa, M., Cresci, G., et al. 2015, A&A, 574, A82, doi: 10.1051/0004-6361/201425035
Pilbratt, G. L., Riedinger, J. R., Passvogel, T., et al. 2010, A&A, 518, L1, doi: 10.1051/0004-6361/201014759
Poglitsch, A., Waelkens, C., Geis, N., et al. 2010, A&A, 518, L2, doi: 10.1051/0004-6361/201014535
Pope, A., Scott, D., Dickinson, M., et al. 2006, MNRAS, 370, 1185, doi: 10.1111/j.1365-2966.2006.10575.x
Rémy-Ruyer, A., Madden, S. C., Galliano, F., et al. 2014, A&A, 563, A31, doi: 10.1051/0004-6361/201322803
Rodighiero, G., Daddi, E., Baronchelli, I., et al. 2011, ApJL, 739, L40, doi: 10.1088/2041-8205/739/2/L40
Rowlands, K., Gomez, H. L., Dunne, L., et al. 2014a, MNRAS, 441, 1040, doi: 10.1093/mnras/stu605
Rowlands, K., Dunne, L., Dye, S., et al. 2014b, MNRAS, 441, 1017, doi: 10.1093/mnras/stu510
Rujopakarn, W., Daddi, E., Rieke, G. H., et al. 2019, arXiv e-prints, arXiv:1904.04507. https://arxiv.org/abs/1904.04507
Saintonge, A., Catinella, B., Tacconi, L. J., et al. 2017, ApJS, 233, 22, doi: 10.3847/1538-4365/aa97e0
Saintonge, A., Wilson, C. D., Xiao, T., et al. 2018, MNRAS, 481, 3497, doi: 10.1093/mnras/sty2499
Salvato, M., Ilbert, O., Hasinger, G., et al. 2011, ApJ, 742, 61, doi: 10.1088/0004-637X/742/2/61
Sanders, D. B., Mazzarella, J. M., Kim, D.-C., Surace, J. A., & Soifer, B. T. 2003, AJ, 126, 1607, doi: 10.1086/376841
Sanders, D. B., Salvato, M., Aussel, H., et al. 2007, ApJS, 172, 86, doi: 10.1086/517885
Santini, P., Maiolino, R., Magnelli, B., et al. 2010, A&A, 518, L154, doi: 10.1051/0004-6361/201014748

Sargent, M. T., Béthermin, M., Daddi, E., & Elbaz, D. 2012, ApJL, 747, L31, doi: 10.1088/2041-8205/747/2/L31
Sargent, M. T., Daddi, E., Béthermin, M., et al. 2014, ApJ, 793, 19, doi: 10.1088/0004-637X/793/1/19
Schinnerer, E., Sargent, M. T., Bondi, M., et al. 2010, ApJS, 188, 384, doi: 10.1088/0067-0049/188/2/384
Schinnerer, E., Groves, B., Sargent, M. T., et al. 2016, ApJ, 833, 112, doi: 10.3847/1538-4357/833/1/112
Schreiber, C., Pannella, M., Elbaz, D., et al. 2015, A&A, 575, A74, doi: 10.1051/0004-6361/201425017
Scoville, N., Aussel, H., Brusa, M., et al. 2007, ApJS, 172, 1, doi: 10.1086/516585
Scoville, N., Aussel, H., Sheth, K., et al. 2014, ApJ, 783, 84, doi: 10.1088/0004-637X/783/2/84
Scoville, N., Sheth, K., Aussel, H., et al. 2016, ApJ, 820, 83, doi: 10.3847/0004-637X/820/2/83
Scoville, N., Lee, N., Vanden Bout, P., et al. 2017, ApJ, 837, 150, doi: 10.3847/1538-4357/aa61a0
Silverman, J. D., Kashino, D., Sanders, D., et al. 2015a, ApJS, 220, 12, doi: 10.1088/0067-0049/220/1/12
Silverman, J. D., Daddi, E., Rodighiero, G., et al. 2015b, ApJL, 812, L23, doi: 10.1088/2041-8205/812/2/L23
—. 2015c, ApJL, 812, L23, doi: 10.1088/2041-8205/812/2/L23
Simpson, J. M., Smail, I., Swinbank, A. M., et al. 2015, ApJ, 807, 128, doi: 10.1088/0004-637X/807/2/128
—. 2017, ApJ, 839, 58, doi: 10.3847/1538-4357/aa65d0
Smith, D. J. B., & Hayward, C. C. 2015, MNRAS, 453, 1597, doi: 10.1093/mnras/stv1727
Smith, D. J. B., Dunne, L., da Cunha, E., et al. 2012, MNRAS, 427, 703, doi: 10.1111/j.1365-2966.2012.21930.x
Smolčić, V., Karim, A., Miettinen, O., et al. 2015, A&A, 576, A127, doi: 10.1051/0004-6361/201424996
Smolčić, V., Novak, M., Bondi, M., et al. 2017, A&A, 602, A1, doi: 10.1051/0004-6361/201628704
Solomon, P. M., & Vanden Bout, P. A. 2005, ARA&A, 43, 677, doi: 10.1146/annurev.astro.43.051804.102221
Song, M., Finkelstein, S. L., Ashby, M. L. N., et al. 2016, ApJ, 825, 5, doi: 10.3847/0004-637X/825/1/5
Speagle, J. S., Steinhardt, C. L., Capak, P. L., & Silverman, J. D. 2014, ApJS, 214, 15, doi: 10.1088/0067-0049/214/2/15
Spilker, J., Bezanson, R., Barišić, I., et al. 2018, ApJ, 860, 103, doi: 10.3847/1538-4357/aac438
Spilker, J. S., Marrone, D. P., Aravena, M., et al. 2016, ApJ, 826, 112, doi: 10.3847/0004-637X/826/2/112
Spreeuw, J. N. 2010, PhD thesis, University of Amsterdam
Swinbank, A. M., Karim, A., Smail, I., et al. 2012, MNRAS, 427, 1066, doi: 10.1111/j.1365-2966.2012.22048.x
Tacconi, L. J., Neri, R., Genzel, R., et al. 2013, ApJ, 768, 74, doi: 10.1088/0004-637X/768/1/74
Tacconi, L. J., Genzel, R., Saintonge, A., et al. 2018, ApJ, 853, 179, doi: 10.3847/1538-4357/aaa4b4
Tan, Q., Daddi, E., Magdis, G., et al. 2014, A&A, 569, A98, doi: 10.1051/0004-6361/201423905
Taniguchi, Y., Scoville, N., Murayama, T., et al. 2007, ApJS, 172, 9, doi: 10.1086/516596
Taniguchi, Y., Kajisawa, M., Kobayashi, M. A. R., et al. 2015, PASJ, 67, 104, doi: 10.1093/pasj/psv106
Tasca, L. A. M., Le Fèvre, O., Ribeiro, B., et al. 2017, A&A, 600, A110, doi: 10.1051/0004-6361/201527963
Umehata, H., Tamura, Y., Kohno, K., et al. 2017, ApJ, 835, 98, doi: 10.3847/1538-4357/835/1/98
Umehata, H., Hatsukade, B., Smail, I., et al. 2018, PASJ, 70, 65, doi: 10.1093/pasj/psy065
Valentino, F., Magdis, G. E., Daddi, E., et al. 2018, ApJ, 869, 27, doi: 10.3847/1538-4357/aaeb88
van der Wel, A., Franx, M., van Dokkum, P. G., et al. 2014, ApJ, 788, 28, doi: 10.1088/0004-637X/788/1/28
van der Wel, A., Noeske, K., Bezanson, R., et al. 2016, ApJS, 223, 29, doi: 10.3847/0067-0049/223/2/29
Walter, F., Brinks, E., de Blok, W. J. G., et al. 2008, AJ, 136, 2563, doi: 10.1088/0004-6256/136/6/2563
Walter, F., Decarli, R., Aravena, M., et al. 2016, ApJ, 833, 67, doi: 10.3847/1538-4357/833/1/67
Wang, T., Elbaz, D., Schreiber, C., et al. 2016, ApJ, 816, 84, doi: 10.3847/0004-637X/816/2/84
Yao, L., Seaquist, E. R., Kuno, N., & Dunne, L. 2003, ApJ, 588, 771, doi: 10.1086/374333
Yun, M. S., Aretxaga, I., Gurwell, M. A., et al. 2015, MNRAS, 454, 3485, doi: 10.1093/mnras/stv1963
Zamojski, M. A., Schiminovich, D., Rich, R. M., et al. 2007, ApJS, 172, 468, doi: 10.1086/516593
Zavala, J. A., Casey, C. M., da Cunha, E., et al. 2018, ApJ, 869, 71, doi: 10.3847/1538-4357/aaecd2
Zhao, Y., Lu, N., Xu, C. K., et al. 2013, ApJL, 765, L13, doi: 10.1088/2041-8205/765/1/L13
—. 2016, ApJ, 819, 69, doi: 10.3847/0004-637X/819/1/69